

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники
 Направление подготовки 09.04.01 Информатика и вычислительная техника
 Отделение школы (НОЦ) Отделение информационных технологий

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

Тема работы
Разработка нейросетевого алгоритма распознавания надписей на изображениях реальных сцен

УДК 004.7.032.26:004.93'12

Студент

Группа	ФИО	Подпись	Дата
8BM92	Лобанова Виктория Александровна		

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Иванова Ю.А.	К.Т.Н.		

КОНСУЛЬТАНТЫ ПО РАЗДЕЛАМ:

По разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Верховская М.В.	К.Э.Н.		

По разделу «Социальная ответственность»

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД ШБИП	Антоневич О.А.	К.б.Н.		

ДОПУСТИТЬ К ЗАЩИТЕ:

Руководитель ООП	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Кочегурова Е. А	К.Т.Н.		

Томск – 2021 г.

ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

по направлению 09.04.01 Информатика и вычислительная техника

Код компетенции	Наименование компетенции
Универсальные компетенции	
УК(У)-1	Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий
УК(У)-2	Способен управлять проектом на всех этапах его жизненного цикла
УК(У)-3	Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели
УК(У)-4	Способен применять современные коммуникативные технологии, в том числе на иностранном (-ых) языке (-ах), для академического и профессионального взаимодействия
УК(У)-5	Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия
УК(У)-6	Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки
Общепрофессиональные компетенции	
ОПК(У)-1	Способен самостоятельно приобретать, развивать и применять математические, естественно-научные, социально-экономические и профессиональные знания для решения нестандартных задач, в том числе в новой или незнакомой среде и в междисциплинарном контексте
ОПК(У)-2	Способен разрабатывать оригинальные алгоритмы и программные средства, в том числе с использованием современных интеллектуальных технологий, для решения профессиональных задач
ОПК(У)-3	Способен анализировать профессиональную информацию, выделять в ней главное, структурировать, оформлять и представлять в виде аналитических обзоров с обоснованными выводами и рекомендациями
ОПК(У)-4	Способен применять на практике новые научные принципы и методы исследований
ОПК(У)-5	Способен разрабатывать и модернизировать программное и аппаратное обеспечение информационных и автоматизированных систем
ОПК(У)-6	Способен разрабатывать компоненты программно-аппаратных комплексов обработки информации и автоматизированного проектирования

ОПК(У)-7	Способен адаптировать зарубежные комплексы обработки информации и автоматизированного проектирования к нуждам отечественных предприятий
ОПК(У)-8	Способен осуществлять эффективное управление разработкой программных средств и проектов
Профессиональные компетенции	
ПК(У)-1	Способен разрабатывать и администрировать системы управления базами данных
ПК(У)-2	Способен проектировать сложные пользовательские интерфейсы
ПК(У)-3	Способен управлять процессами и проектами по созданию (модификации) информационных ресурсов
ПК(У)-4	Способен осуществлять руководство разработкой комплексных проектов на всех стадиях и этапах выполнения работ
ПК(У)-5	Способен проектировать и организовывать учебный процесс по образовательным программам с использованием современных образовательных технологий

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники
Направление подготовки 09.04.01 Информатика и вычислительная техника
Отделение школы (НОЦ) Отделение информационных технологий

УТВЕРЖДАЮ:
Руководитель ООП
_____ Кочегурова Е.А.
(Подпись) (Дата) (Ф.И.О.)

ЗАДАНИЕ на выполнение выпускной квалификационной работы

В форме:

Магистерской диссертации
(бакалаврской работы, дипломного проекта/работы, магистерской диссертации)

Студенту:

Группа	ФИО
8BM92	Лобановой Виктории Александровне

Тема работы:

Разработка нейросетевого алгоритма распознавания надписей на изображениях реальных сцен	
Утверждена приказом директора (дата, номер)	№ 95-29/с от 05.04.2021

Срок сдачи студентом выполненной работы:	01.06.2021
--	------------

ТЕХНИЧЕСКОЕ ЗАДАНИЕ:

<p>Исходные данные к работе <i>(наименование объекта исследования или проектирования; производительность или нагрузка; режим работы (непрерывный, периодический, циклический и т. д.); вид сырья или материал изделия; требования к продукту, изделию или процессу; особые требования к особенностям функционирования (эксплуатации) объекта или изделия в плане безопасности эксплуатации, влияния на окружающую среду, энергозатратам; экономический анализ и т. д.).</i></p>	<p>Объектом проектирования и разработки является нейросетевой алгоритм распознавания надписей на изображениях реальных сцен.</p>
<p>Перечень подлежащих исследованию, проектированию и разработке вопросов <i>(аналитический обзор по литературным источникам с целью выяснения достижений мировой науки техники в рассматриваемой области; постановка задачи исследования, проектирования, конструирования; содержание процедуры исследования, проектирования, конструирования; обсуждение результатов выполненной работы; наименование дополнительных разделов, подлежащих разработке; заключение по работе).</i></p>	<ol style="list-style-type: none"> 1. Аналитический обзор существующих методов обнаружения текстовых областей на изображениях реальных сцен; 2. Разработка алгоритма для распознавания текстовых областей; 3. Проектирование и реализация программного обеспечения на основании алгоритма; 4. Работа над разделом по финансовому менеджменту, ресурсоэффективности и ресурсосбережения;

	5. Работа над разделом по социальной ответственности; 6. Работа над разделом на английском языке.
Перечень графического материала (с точным указанием обязательных чертежей)	1. Алгоритм работы программы 2. Схемы структур сверточных нейронных сетей 3. Графики сравнительных обучений сетей 4. Таблицы сравнительных обучений сетей 5. Предсказанные маски для тестовых выборок
Консультанты по разделам выпускной квалификационной работы (с указанием разделов)	
Раздел	Консультант
Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	Доцент ОСГН ШБИП, к.э.н., Верховская М.В.
Социальная ответственность	Доцент ООД ШБИП, к.б.н., Антоневиц О.А.
Английский язык	Доцент ОИЯ ШБИП, к.ф.н., Айкина Т.Ю.
Названия разделов, которые должны быть написаны на русском и иностранном языках:	
Раздел 1 Text detection methods	

Дата выдачи задания на выполнение выпускной квалификационной работы по линейному графику	01.03.2021
---	------------

Задание выдал руководитель:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Иванова Ю.А.	к.т.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8BM92	Лобанова Виктория Александровна		

Министерство науки и высшего образования Российской Федерации
 федеральное государственное автономное
 образовательное учреждение высшего образования
 «Национальный исследовательский Томский политехнический университет» (ТПУ)

Школа Инженерная школа информационных технологий и робототехники

Направление подготовки 09.04.01 Информатика и вычислительная техника

Уровень образования Магистратура

Отделение школы (НОЦ) Отделение информационных технологий

Период выполнения _____ (осенний / весенний семестр 2020 /2021 учебного года)

Форма представления работы:

Магистерская диссертация
(бакалаврская работа, дипломный проект/работа, магистерская диссертация)

КАЛЕНДАРНЫЙ РЕЙТИНГ-ПЛАН выполнения выпускной квалификационной работы

Срок сдачи студентом выполненной работы:	01.06.2021
--	------------

Дата контроля	Название раздела (модуля) / вид работы (исследования)	Максимальный балл раздела (модуля)
01.06.2021	Анализ методов распознавания текстовых областей	20
01.06.2021	Проектирование алгоритма	20
01.06.2021	Реализация программного обеспечения на основании алгоритма	30
01.06.2021	Финансовый менеджмент, ресурсоэффективность и ресурсосбережение	10
01.06.2021	Социальная ответственность	10
01.06.2021	Приложение на английском языке	10

СОСТАВИЛ:

Руководитель ВКР

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Иванова Ю.А.	К.Т.Н.		

СОГЛАСОВАНО:

Руководитель ООП

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Кочегурова Е. А	К.Т.Н.		

**ЗАДАНИЕ ДЛЯ РАЗДЕЛА
«ФИНАНСОВЫЙ МЕНЕДЖМЕНТ, РЕСУРСОЭФФЕКТИВНОСТЬ И
РЕСУРСОСБЕРЕЖЕНИЕ»**

Студенту:

Группа	ФИО
8BM92	Лобановой Виктории Александровне

Школа	ИШИТР	Отделение школы (НОЦ)	ОИТ
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

Исходные данные к разделу «Финансовый менеджмент, ресурсоэффективность и ресурсосбережение»:

1. Стоимость ресурсов научного исследования (НИ): материально-технических, энергетических, финансовых, информационных и человеческих	Оклад инженера – 22695 руб., Оклад научного руководителя – 35111,5 руб.
2. Нормы и нормативы расходования ресурсов	Районный коэффициент 30%; Коэффициент дополнительной заработной платы 12%; Накладные расходы 16%.
3. Используемая система налогообложения, ставки налогов, отчислений, дисконтирования и кредитования	Коэффициент отчислений на уплату во внебюджетные фонды 27,1 %.

Перечень вопросов, подлежащих исследованию, проектированию и разработке:

1. Оценка коммерческого потенциала, перспективности и альтернатив проведения НИ с позиции ресурсоэффективности и ресурсосбережения	Анализ конкурентных технических решений. SWOT-анализ.
2. Планирование и формирование бюджета научных исследований	Формирование плана и графика разработки: - определение структуры и трудоемкости работ, - создание диаграммы Ганта. Формирование бюджета затрат на разработку: - затраты на амортизацию оборудования; - заработная плата (основная и дополнительная), - отчисления во внебюджетные фонды, - накладные расходы.
3. Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности разработки	Определение потенциального эффекта разработки.

Перечень графического материала (с точным указанием обязательных чертежей):

1. Оценочная карта конкурентных технических решений, 2. Матрица SWOT-анализа и интерактивная матрица проекта, 3. График проведения НИИ, 4. Диаграмма Ганта, 5. Бюджет НИИ, 6. Оценка ресурсной, финансовой и экономической эффективности НИИ

Дата выдачи задания для раздела по линейному графику	01.03.2021
--	------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОСГН ШБИП	Верховская М.В.	К.Э.Н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8BM92	Лобанова Виктория Александровна		

ЗАДАНИЕ ДЛЯ РАЗДЕЛА «СОЦИАЛЬНАЯ ОТВЕТСТВЕННОСТЬ»

Студенту:

Группа	ФИО
8BM92	Лобановой Виктории Александровне

Школа	ИШИТР	Отделение (НОЦ)	ОИТ
Уровень образования	Магистратура	Направление/специальность	09.04.01 Информатика и вычислительная техника

Тема ВКР:

Разработка нейросетевого алгоритма распознавания надписей на изображениях реальных сцен	
Исходные данные к разделу «Социальная ответственность»:	
1. Характеристика объекта исследования (вещество, материал, прибор, алгоритм, методика, рабочая зона) и области его применения	<p>Объект исследования: нейросетевой алгоритм</p> <p>Область применения: распознавание надписей на изображениях реальных сцен</p> <p>Рабочая зона: рабочее место для выполнения работ сидя с использованием ПЭВМ</p>
Перечень вопросов, подлежащих исследованию, проектированию и разработке:	
1. Правовые и организационные вопросы обеспечения безопасности: <ul style="list-style-type: none"> – специальные (характерные при эксплуатации объекта исследования, проектируемой рабочей зоны) правовые нормы трудового законодательства; – организационные мероприятия при компоновке рабочей зоны. 	<ul style="list-style-type: none"> – ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования – СП 2.4.3648-20 «Санитарно-эпидемиологические требования к организациям воспитания и обучения, отдыха и оздоровления детей и молодежи»
2. Производственная безопасность: 2.1. Анализ выявленных вредных и опасных факторов 2.2. Обоснование мероприятий по снижению воздействия	<p>Вредные факторы:</p> <ul style="list-style-type: none"> – Отсутствие или недостаток необходимого искусственного освещения – Отклонение показателей микроклимата в закрытом помещении – Повышенный уровень электромагнитных излучений – Повышенный уровень шума на рабочем месте – Умственное перенапряжение – Перенапряжение зрительного анализатора – Статические физические нагрузки, связанные с рабочей позой <p>Опасные факторы:</p> <ul style="list-style-type: none"> – Повышенное значение напряжения
3. Экологическая безопасность:	Литосфера: бытовой мусор и компьютерная техника, подлежащая утилизации
4. Безопасность в чрезвычайных ситуациях:	Наиболее типичная ЧС: пожар

Дата выдачи задания для раздела по линейному графику	01.03.2021
--	------------

Задание выдал консультант:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ООД ШБИП	Антоневич О.А.	к.б.н.		

Задание принял к исполнению студент:

Группа	ФИО	Подпись	Дата
8BM92	Лобанова Виктория Александровна		

Реферат

Выпускная квалификационная работа состоит из 103 с., 39 рис., 25 табл., 39 источников.

Ключевые слова: разработка, сверточная нейронная сеть, машинное обучение, обработка изображений, текстовые области.

Объектом проектирования и разработки является нейросетевой алгоритм распознавания надписей на изображениях реальных сцен.

Целью работы является разработка алгоритма распознавания текстовых областей на изображениях реальных сцен. Для её достижения необходимо выполнение следующих задач: обзор существующих моделей, выбор модели для реализации, разработка алгоритма детектирования текстовых областей на изображениях, проектирование программного обеспечения на его основании, реализация согласно разработанному проекту и тестирование.

В процессе исследования был проведён анализ существующих моделей обнаружения текстовых областей на изображениях реальных сцен, на основании которого был произведён выбор модели для дальнейшей реализации.

В результате был предложен нейросетевой алгоритм распознавания надписей на изображениях реальных сцен. В ходе проведения тестирования были подобраны такие параметры нейронной сети как размеры входных изображений, количество и типы используемых слоёв. Было проведено увеличение исходной базы изображений за счёт применения поворотов, сжатия и разделения входящих в неё изображений. В результате конечного обучения нейросетевого алгоритма удалось добиться высокой точности: 91,1% для обучающей выборки, 90,0% для валидационной и 88,0% для тестовой.

Было проведено сравнение с работами, проводившими тестирование на аналогичной базе изображений, в ходе которого предложенный алгоритм показал наилучшие результаты.

Содержание

Введение.....	13
1 Методы распознавания текстовых областей.....	14
1.1 Методы связанных компонент.....	14
1.2 Текстурные методы.....	16
1.3 Методы глубокого обучения.....	20
1.4 Сравнение методов детектирования текстовых областей на изображениях	23
2 Проектирование алгоритма.....	26
2.1 Используемые инструменты.....	26
2.2 Модули программы для обучения сверточной нейронной сети.....	28
2.3 Архитектура программного обеспечения.....	28
2.4 Описание структуры сети.....	29
2.5 Параметры оценки при обучении сети.....	31
2.6 Фильтры для предобработки изображений.....	32
2.7 Преобразование Фурье.....	33
2.8 Сглаживающие частотные фильтры.....	33
3 Результаты обучения сети.....	35
3.1 Подбор размеров входных изображений.....	35
3.2 Подбор количества слоёв.....	38
3.3 Подбор типов свёрточных слоёв.....	39
3.4 Предобработка входных изображений.....	40
2.5 Частотный анализ с использованием преобразования Фурье.....	45
3.7 Сравнение с другими работами.....	54
4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение.....	58
4.1 Предпроектный анализ.....	58
4.2 Планирование научно-исследовательских работ.....	61
4.3 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования.....	66
4.4 Вывод по разделу.....	69

5 Социальная ответственность	70
5.1 Правовые и организационные вопросы обеспечения безопасности	70
5.2 Профессиональная социальная безопасность.	72
5.3. Экологическая безопасность.....	82
5.4. Безопасность в чрезвычайных ситуациях.....	83
5.5 Выводы по разделу.....	84
Заключение	86
Список использованных источников	87
Приложение А	91
Приложение Б	102

Введение

В настоящее время существует огромное количество информации, хранящейся в виде изображений, содержание которых представляет собой определенную ценность. Детектирование и последующее распознавание текста на изображениях может быть применено в таких областях как перевод фотографий документов в текстовую форму [1], автоматическое определение номерных знаков автомобилей [2], геолокация объекта по названиям улиц, улучшение качества детектирования и распознавания объектов на изображениях. Однако объемы информации, хранящейся в виде изображений, велики, что делает невозможным ее обработку вручную. Тем не менее, автоматизированные методы обработки изображений позволяют успешно справляться с этой задачей.

Несмотря на широкую область возможного применения и прогресс в сфере машинного обучения, обнаружение и сегментация текстовых областей на изображениях все еще представляет собой проблему. Текст, расположенный на изображениях реальных сцен, может быть различным по размеру, стилю, цвету, повороту относительно горизонта. Также возможно его перекрытие другими объектами на изображении или низкая контрастность с фоном. Применение фильтров для сглаживания шумов и увеличения контрастности на границах объектов не является универсальным решением проблемы, так как может, как вызывать ложные срабатывания, так и мешать обнаружению значимых областей.

Целью данной работы является разработка нейросетевого алгоритма распознавания надписей на изображениях реальных сцен.

Для достижения данной цели необходимо:

- проанализировать существующие модели обнаружения текстовых областей на изображениях реальных сцен;
- выбрать модель для дальнейшей реализации;
- разработать алгоритм для распознавания надписей;
- реализовать программу на основе разработанного алгоритма;
- произвести обучение и тестирование разработанного алгоритма.

1 Методы распознавания текстовых областей

Методы распознавания текстовых областей используются для определения наличия и выделения местоположения текстовых областей на изображениях. Однако, точность данного определения может быть снижена из-за различных размеров, стилей и направлений надписей. Кроме того, низкий контраст с фоном или сложные фоны могут вызывать дополнительные затруднения. Все существующие методы распознавания текстовых областей можно разделить на следующие группы:

- методы связанных компонент;
- текстурные методы;
- методы глубокого обучения.

1.1 Методы связанных компонент

Принцип работы методов связанных компонент заключается в поиске и объединении малых компонент в большие на основании определённых характеристик пикселей: яркость, цвет, толщина контура элемента. Затем из полученных компонент извлекаются признаки для дальнейшей классификации на текстовые и нетекстовые компоненты. Найденные текстовые компоненты извлекаются из изображений и объединяются в текстовые области.

Главными преимуществами методов связанных компонент являются простота вычислений и высокая точность. Однако данные методы плохо работают при изменении поворота или масштаба изображений, обработке сложных фонов и некоторых других случаях.

Для рассмотрения были выбраны метод максимально стабильных экстремальных областей (MSERs) и метод преобразования по толщине штриха символа текста (SWT).

1.1.1 Метод максимально стабильных экстремальных областей

Метод MSERs использует экстремальные области для распознавания текстовых областей. Экстремальные области на изображении определяются двумя условиями относительно множества пикселей внутри области [3]:

- область остаётся замкнутой при преобразовании координат (поворот, растяжение);
- область остаётся замкнутой при изменении яркости изображения.

Применение метода MSERs для различных пороговых значений представлено на рисунке 1.1.

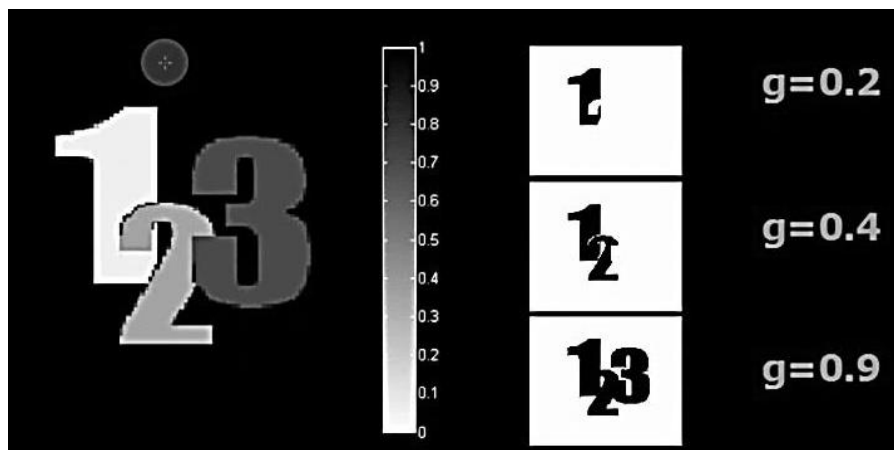


Рисунок 1.1 – Пример применения метода MSERs

Данные методы требуют малое количество памяти для реализации, могут работать в режиме реального времени, но плохо распознают текстовые области на размытых или неконтрастных изображениях [4].

1.1.2 Метод преобразования по толщине штриха символа текста

Метод SWT строится на предположении о том, что буквы и символы на текстовых областях имеют определенные геометрические особенности: одинаковая ширина для каждого символа и одинаковая толщина штриха символа текста. Другими словами, в большинстве случаев текстовые области имеют небольшое изменение толщины штриха от символа к символу, в отличие от нетекстовых областей [5].

Метод SWT преобразует изображение в массив, где каждый элемент содержит значение толщины штриха для соответствующего пикселя, используя для этого оператор Кэнни. Далее пиксели группируются в предполагаемые текстовые области. Два соседних пикселя объединяются, если они обладают равными значениями толщины штриха. Отделить текстовые области от нетекстовых позволяет одинаковое значение толщины для текста. Затем из выделенных текстовых областей формируется предсказанное текстовое поле.

На рисунке 1.2 показана последовательность применения метода SWT.

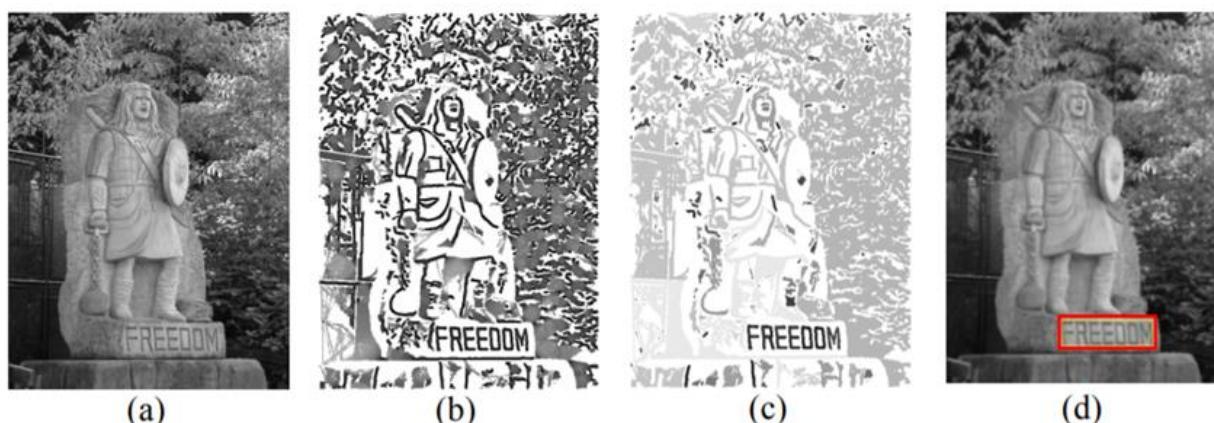


Рисунок 1.2 - Пример применения метода SWT (a – исходное изображение, b – изображение со значениями ширины штриха для каждого пикселя, c – изображение с выделенными текстовыми областями, d – изображение с предсказанным текстовым полем)

Эксперименты показывают, что метод SWT очень эффективен для обнаружения текста. Данный метод может быть применён к различным шрифтам и языкам, а также к различным размерам и поворотам надписей. Однако, метод SWT не является полностью автоматическим и требует предварительного подбора параметров, что может привести к ложным срабатываниям для сложных случаев.

1.2 Текстурные методы

Текстурные методы основаны на идее о том, что текстовые области отличаются от фона изображений высокочастотными и регулярными текстурными признаками. В данных методах происходит применение классификатора к отдельным областям изображения для определения наличия или отсутствия текста. Текстурные методы хорошо показывают себя на зашумленных и искажённых изображениях, а также при наличии сложного текста или фона. Однако, такая классификация сложна в вычислительном отношении, что сказывается на времени работы методов. Дополнительные трудности могут быть вызваны тем, что наравне с текстовыми областями на изображениях улиц регулярностью обладают такие объекты, как кирпичная кладка, окна домов, листва деревьев.

Для рассмотрения были выбраны следующие текстурные методы:

- дискретно-косинусное преобразование;
- гистограмма направленных градиентов;
- признаки Хаара.

1.2.1 Дискретно-косинусное преобразование

Дискретно-косинусное преобразование было создано для сжатия изображений, которое позволяет трансформировать пространство изображения в пространство свойств с более низкой размерностью [6].

Однако данный метод можно использовать и для извлечения текстовых областей из изображений. Преобразование применяется для каждого отдельного фрагмента изображения независимо друг от друга, после чего происходит объединение значений. Значения полученных частотных коэффициентов отражают локальную периодичность в области изображения. Неявная периодичность означает, что на границах возникают разрывы.

Алгоритм, предложенный Zhong, использует дискретно-косинусное преобразование и состоит из двух основных этапов [7]:

- обнаружение предполагаемых текстовых областей в сжатой области частотных коэффициентов;
- постобработка областей.

Применение дискретно-косинусного преобразования показано на рисунке 1.3.

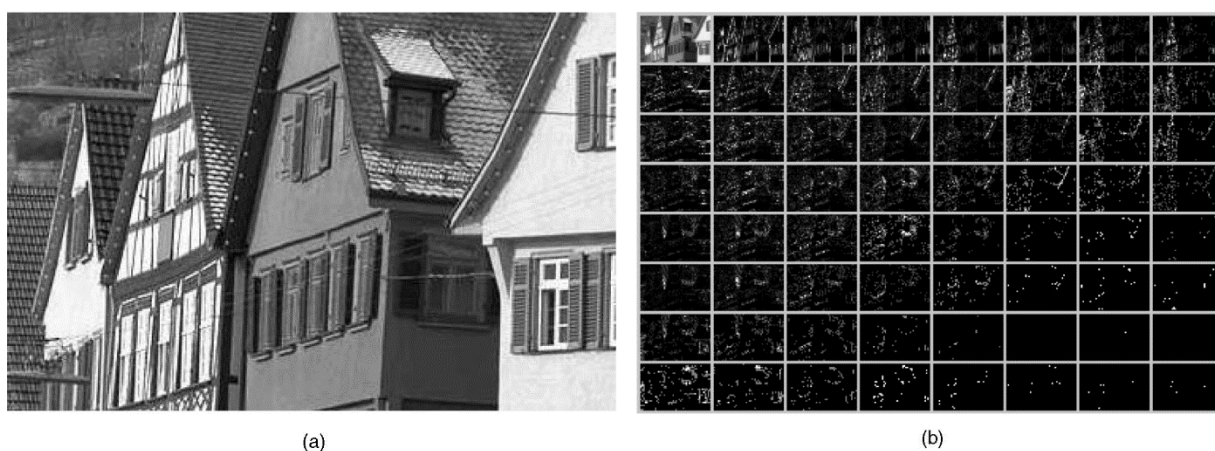


Рисунок 1.3 - Применение дискретно-косинусного преобразования (а – исходное изображение, б – изображения с выделенными частотными коэффициентами)

Предложенный алгоритм обладает высокой скоростью за счёт работы с пространством свойств с более низкой размерностью и постобработкой изображений с пониженным разрешением. Однако данный алгоритм не предоставляет высокой точности при разделении областей на текстовые и нетекстовые.

1.2.2 Гистограмма направленных градиентов

Ещё одним рассматриваемым текстурным методом является использование HOG-дескрипторов или применение метода гистограмм направленных градиентов (Histogram of Oriented Gradients). Сначала происходит разделение изображения на ячейки, затем для каждого пикселя внутри ячейки рассчитывается гистограмма направлений градиентов. Объединение полученных значений для каждой ячейки называется HOG-дескриптором. Полученное градиентное изображение выделяет контуры на изображении и исключает из рассмотрения несущественную информацию [8].

На рисунке 1.4 показан пример применения данного метода для распознавания цифр. Для каждой цифры представлены гистограммы, показывающие преобладание каждого градиента [9].

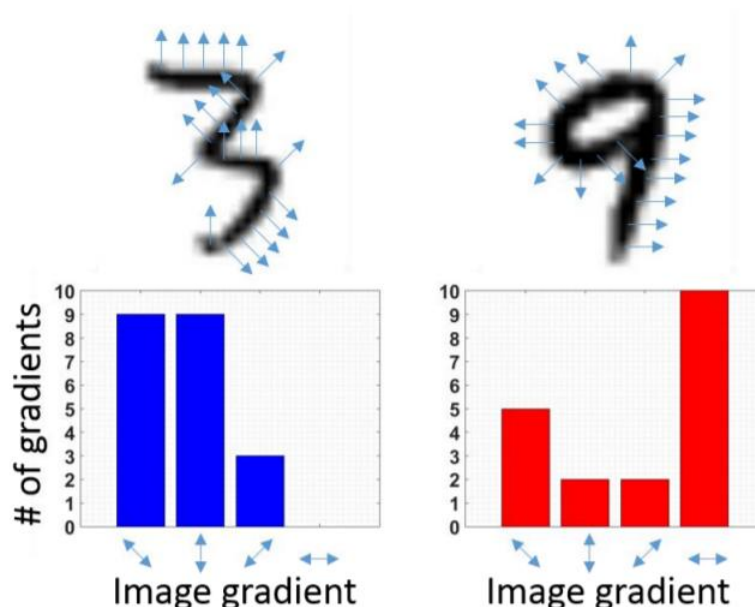


Рисунок 1.4 - пример применения метода гистограмм направленных градиентов

В данном примере рассматриваются отдельные цифры, а не полноценный текст. Такие данные легко визуализировать и интерпретировать человеком в

отличии от изображений реальных сцен, на которых помимо текста присутствуют и другие объекты. Поэтому метод гистограмм направленных градиентов больше подходит для предварительной обработки изображений, но не для самостоятельного применения.

1.2.3 Признаки Хаара

Признаки Хаара используют вейвлет-преобразование Хаара. Данный метод использует в качестве признаков прямоугольные области, разделенные на контрастные части. Данные признаки применяются путём наложения на фрагменты рассматриваемого изображения, при наложении происходит вычисление суммы интенсивностей пикселей для каждой области признака. Различие между полученными значениями позволяет распределить по категориям фрагменты изображения [10].

В большинстве случаев текст представляет собой несколько хорошо контрастирующих по горизонтали или вертикали частей на изображении независимо от цвета текста или фона. Данный факт предоставляет возможность для применения признаков Хаара для обнаружения текстовых областей [11-12].

На рисунке 1.5 представлены примеры из набора признаков Хаара.

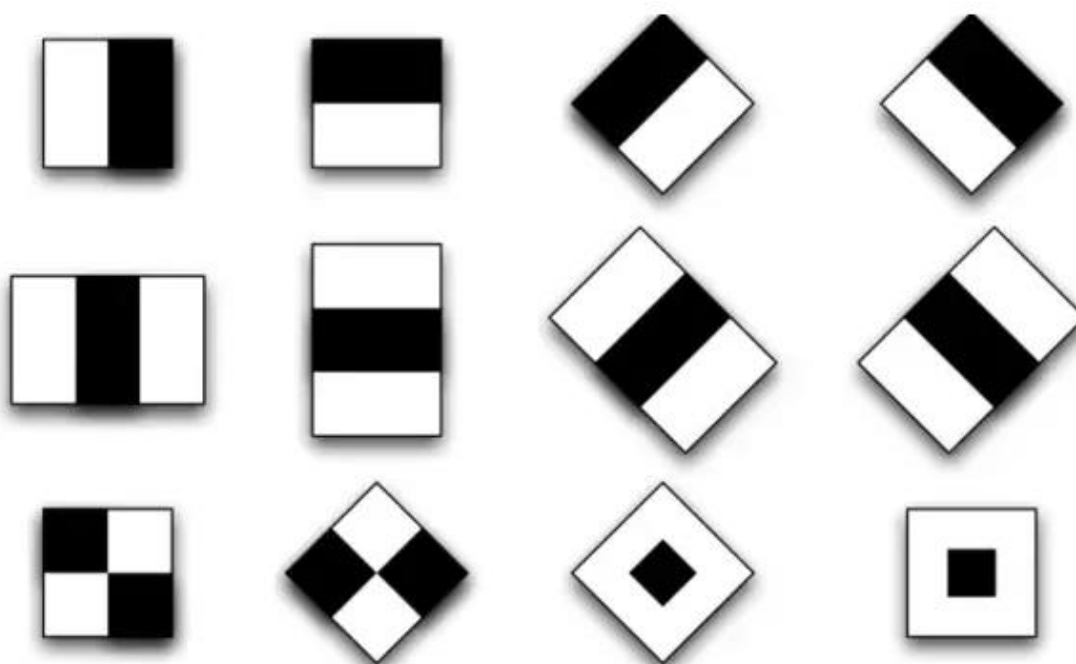


Рисунок 1.5 - Признаки Хаара

Главным минусом применения данного метода для обнаружения текстовых областей является зависимость результата от начального выбора набора применяемых признаков. Позиции букв в тексте и их формы могут сильно различаться для разных изображений, что может вызвать дополнительную проблему при подготовке набора признаков.

1.3 Методы глубокого обучения

Методы глубокого обучения являются областью машинного обучения; наиболее распространёнными методами являются различные типы нейронных сетей.

Полносвязная нейронная сеть состоит из слоев нейронов. Нейроны из каждого предыдущего слоя связаны с нейронами последующего слоя. В процессе обучения и работы сети нейроны последовательно активируются слой за слоем. Процесс активации представляет собой формирование выходного сигнала нейрона в ответ на полученный входной сигнал. Обучение такой нейронной сети заключается в подборе коэффициентов весов связей между нейронами таким образом, чтобы выходные значения сети соответствовали эталонным [13].

1.3.1 Сверточные нейронные сети

Сверточные нейронные сети (CNN) являются классом нейронных сетей. В отличие от полносвязных нейронных сетей, где каждый нейрон предыдущего слоя связан со всеми нейронами следующего слоя, в сверточных нейронных сетях соседние нейроны одного слоя связаны с локальной областью нейронов последующего слоя, что значительно сокращает количество настраиваемых во время обучения весов.

Это делает сверточные сети более вычислительноэффективными, позволяя использовать в качестве входных значений изображения большего размера. Сверточные сети также могут включать подвыборочные слои, используемые для уменьшения размеров данных за счёт объединения выходов кластеров нейронов в один входной нейрон [14].

Структура сверточной нейронной сети представлена на рисунке 1.6.

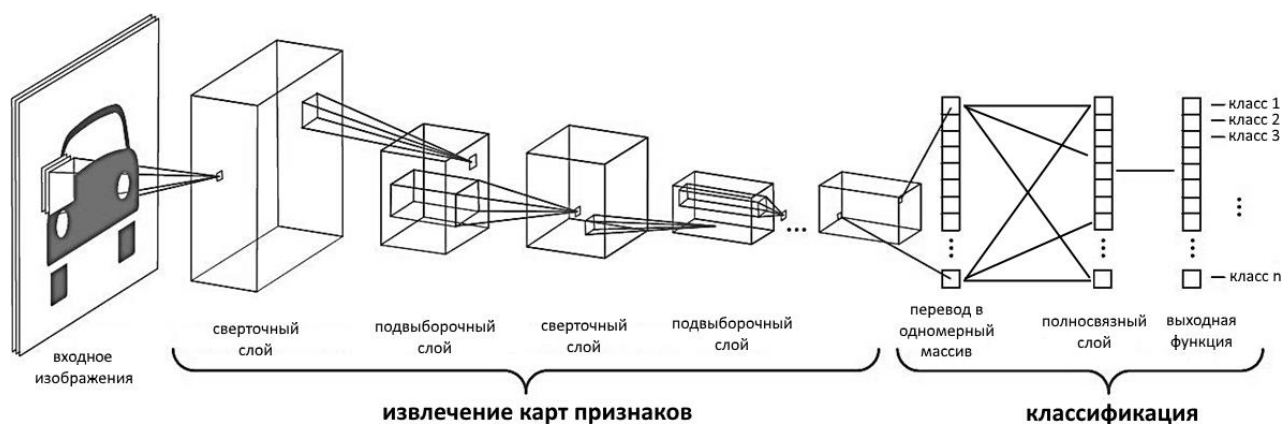


Рисунок 1.6 - Структура сверточной нейронной сети

Основным преимуществом сверточной сети является возможность работы со сложными изображениями, так как она обладает инвариантностью к поворотам, растяжениям, изменениям яркости или контрастности [14]. Такая сеть может извлекать информацию из изображения с меньшими по сравнению с полносвязными сетями вычислительными затратами. Недостатком сверточной нейронной сети являются её выходные значения, представляемые в виде вектора значений: данная сеть может быть использована для обнаружения наличия текста на изображении, но не для определения его местоположения.

1.3.2 Полносвёрточные нейронные сети

Структура полносвёрточных нейронных сетей (FCN) схожа со структурой свёрточной нейронной сети. Главное отличие заключается в выходном слое сети. Данный слой производит объединение всех карт признаков с предыдущих слоёв для создания изображения в виде сегментированной карты, каждый сегмент которой представляет собой область определенного класса. Полученная сегментированная карта может быть использована для определения положения текстовых областей на изображении [15].

Пример структуры полносвёрточной нейронной сети показан на рисунке 1.7.

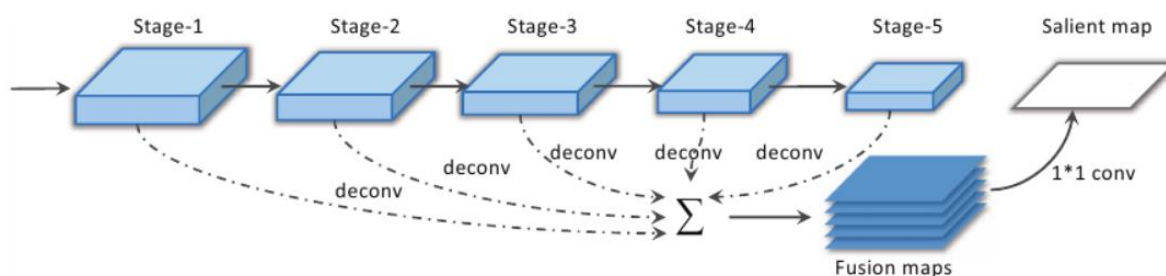


Рисунок 1.7 - Пример структуры полносвёрточной нейронной сети

При использовании полносвёрточной нейронной сети могут появляться ложные срабатывания или пропущенные символы при очень низкой контрастности, кривизне, сильном отражении света, плотном расположении строк текста или больших промежутках между символами. Еще одним недостатком данного типа сети является быстроедействие.

1.3.3 Архитектура U-Net

Выходное изображение полносверточной нейронной сети архитектуры U-Net совпадает по размерам со входным изображением. Отличительной особенностью U-Net является ее симметричная структура. Она состоит из сужающей и расширяющей частей, которые создают U-образную структуру. Расширяющая часть состоит из слоев повышающей дискретизации, которые позволяют обрабатывать большее количество карт признаков. В данной структуре отсутствуют полносвязные слои, в результате выходное изображение сети содержит предсказания только для тех пикселей, которые соответствуют входному изображению [16].

Пример нейронной сети архитектуры U-Net представлен на рисунке 1.8.

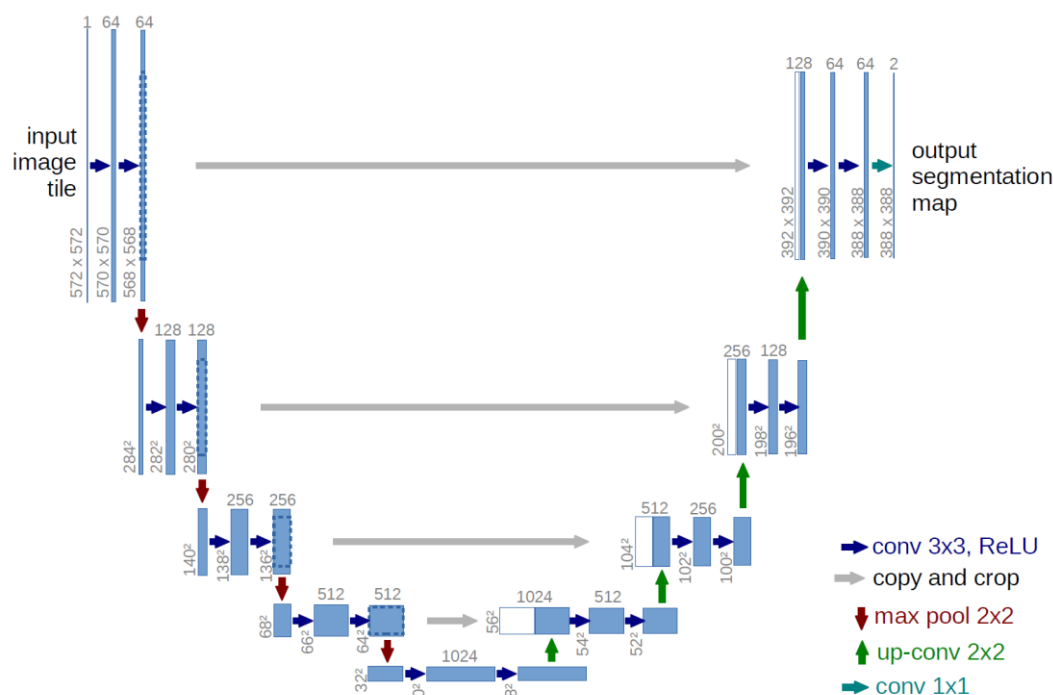


Рисунок 1.8 - Пример нейронной сети архитектуры U-Net

Основным преимуществом архитектуры U-Net является наличие большего количества слоев повышающей дискретизации, которые позволяют обрабатывать большее количество карт признаков. Кроме того, U-Net обеспечивает попиксельную точность благодаря одинаковому размеру входных и выходных изображений [17].

1.4 Сравнение методов детектирования текстовых областей на изображениях

Сравнение рассмотренных методов детектирования текстовых областей на изображениях представлено в таблице 1.1. Для сравнения были рассмотрены следующие характеристики:

- Скорость работы (количество времени, необходимое для обучения данного метода);
- Ресурсная стоимость вычислений (требования к вычислительным ресурсам оборудования);
- Работа со сложными изображениями:
 - Геометрические преобразования (повороты, растяжения);
 - Текстурные искажения (шум, низкая контрастность, сложные фоны);

- Эффективность метода (точность или другие меры оценки).

Таблица 1.1 - Сравнение методов

Характеристика Метод		Скорость работы	Ресурсная стоимость вычислений	Геометрические преобразования	Текстурные искажения	Эффективность метода
Методы связанных компонент	Метод максимально стабильных экстремальных областей	Высокая	Низкая	Не работает	Не работает	Средняя
	Метод преобразования по толщине штриха символа текста	Средняя	Низкая	Работает	Не работает	Средняя
Текстурные методы	Дискретно-косинусное преобразование	Высокая	Низкая	Не работает	Не работает	Низкая
	HOG-дескрипторы	Средняя	Низкая	Не работает	Не работает	Низкая
	Признаки Хаара	Средняя	Низкая	Не работает	Не работает	Низкая
Методы глубокого обучения	Свёрточная нейронная сеть	Средняя	Низкая	Работает	Работает	Средняя
	Полносвёрточная нейронная сеть	Низкая	Средняя	Работает	Работает	Высокая
	Сеть архитектуры U-Net	Низкая	Средняя	Работает	Работает	Высокая

Как видно из таблицы, методы связанных компонент обладают высокой скоростью работы, однако плохо работают с различными искажениями изображений и имеют среднюю эффективность метода. Текстурные методы обладают средней скоростью работы, плохо работают с различными искажениями и имеют низкую эффективность. Методы глубокого обучения являются самой медленной группой из-за длительного времени обучения, но хорошо работают с различными искажениями изображений, а также обладают высокой производительностью. Следовательно, методы, основанные на глубоком обучении, представляют большую ценность для рассмотрения.

Свёрточные нейронные сети предъявляют меньше требований к оборудованию по сравнению с другими сетями, но являются недостаточными

для определения местоположения текстовых областей на изображениях. Полносвёрточные нейронные сети с несимметричной структурой и сети архитектуры U-Net находятся на одном уровне. Однако симметричная архитектура сети U-Net предоставляет больше возможностей для распознавания текстовых областей ввиду наличия большего количества слоев повышающей дискретизации для обработки карт признаков. По этой причине архитектура U-Net была выбрана для реализации алгоритма в данной работе.

2 Проектирование алгоритма

2.1 Используемые инструменты

Для реализации программного обеспечения был выбран высокоуровневый язык программирования Python, так как для него реализованы библиотеки для работы со сверточными нейронными сетями. Были использованы функции открытой нейросетевой библиотеки Keras, являющейся надстройкой над библиотекой машинного обучения TensorFlow. Также для работы с многомерными массивами изображений было задействовано расширение языка NumPy. Для обработки изображений была задействована библиотека scikit-image.

Написание программного кода производилось в интегрированной среде разработки PyCharm. Данная среда разработки включает в себя интегрированный отладчик и запуск тестирования, предоставляет удобный интерфейс для работы с программным кодом.

Для обучения сверточной нейронной сети был использован набор облачных служб Google Cloud Platform, предоставляющий такие службы как облачные вычисления, хранение данных, анализ данных и машинное обучение. Была создана виртуальная машина для обучения сверточных нейронных сетей.

В качестве базы изображений была выбрана база KAIST Scene Text Database, из которой было выбрано 1215 фотографий различных вывесок, содержащих текст на английском и корейском языке. Изображения были получены с помощью цифровой камеры высокого разрешения или мобильного телефона с низким разрешением. Все изображения были приведены к размеру 640×480 пикселя.

Для определения расположения текста для каждого изображения представлено изображение-маска, на котором черным цветом выделен фон, а красным – текст.

Примеры изображений из базы изображений KAIST Scene Text Database представлены на рисунке 2.1.



Рисунок 2.1 - Примеры изображений из базы изображений KAIST Scene Text Database

Было проведено искусственное увеличение базы изображений в 7 раз за счёт поворота и обрезки изображений. Повороты производились в обе стороны от -18° до 18° с шагом в 6° . Также все маски были переведены в чёрно-белый формат. Результат такой обработки для одного изображения представлен на рисунке 2.2.



Рисунок 2.2 - Поворот одного изображения из базы изображений

Дополнительный способ увеличения базы изображений представлен на рисунке 2.3, где к одному изображению применяется как сжатие до размеров, используемых для обучаемой структуры сети, так и разделение изображения на 4 части с перекрытием.



Рисунок 2.3 - Сжатие и разделение с перекрытием одного изображения из базы изображений

2.2 Модули программы для обучения сверточной нейронной сети

Программа состоит из двух модулей:

- модуль предобработки базы изображений;
- модуль, ответственного за работу со сверточной нейронной сетью.

При запуске работы модуля предобработки базы изображений сначала происходит загрузка данных, применение фильтров к входным изображениям, а также загрузка изображений-масок, с которыми далее будет происходить сравнение выходов обучаемой сети. Для базы изображений формируются пр-массивы, к которым в дальнейшем будет происходить обращение.

При запуске работы модуля, ответственного за работу со сверточной нейронной сетью, считываются пр-массивы базы данных изображений. База изображений делится на обучающую, валидационную и тестовую выборки в соотношении 72%, 18% и 10% соответственно. Далее происходит создание нейронной сети с заранее прописанными параметрами слоев. После этого возможен вызов обучения созданной нейронной сети на переданных классу выборках. После завершения обучения производится сохранение параметров сети. Затем сеть запускается для предсказаний на тестовых значениях, каждое из которых сохраняется в отдельный файл.

2.3 Архитектура программного обеспечения

При запуске программы предобработки изображений из базы данных происходит загрузка изображений их предварительная фильтрация и загрузка изображений-масок.

При запуске разработанного программного обеспечения, связанной с нейронной сетью, происходит разделение ранее загруженной и обработанной базы изображений на обучающую, валидационную и тестовую выборки в соотношении 72%, 18% и 10% соответственно. Далее создается нейронная сеть, происходит ее обучение, валидация и тестирование на сформированных ранее выборках. После окончания обучения программа сохраняет параметры сверточной нейронной сети.

Объединённый алгоритм работы программ в виде блок-схемы представлен на рисунке 2.4.



Рисунок 2.4 - Объединённый алгоритм работы программ

2.4 Описание структуры сети

Для обучения используются 2 конфигурации сети: первая представлена на рисунке 2.5 и используется для конечного обучения сети U-Net.

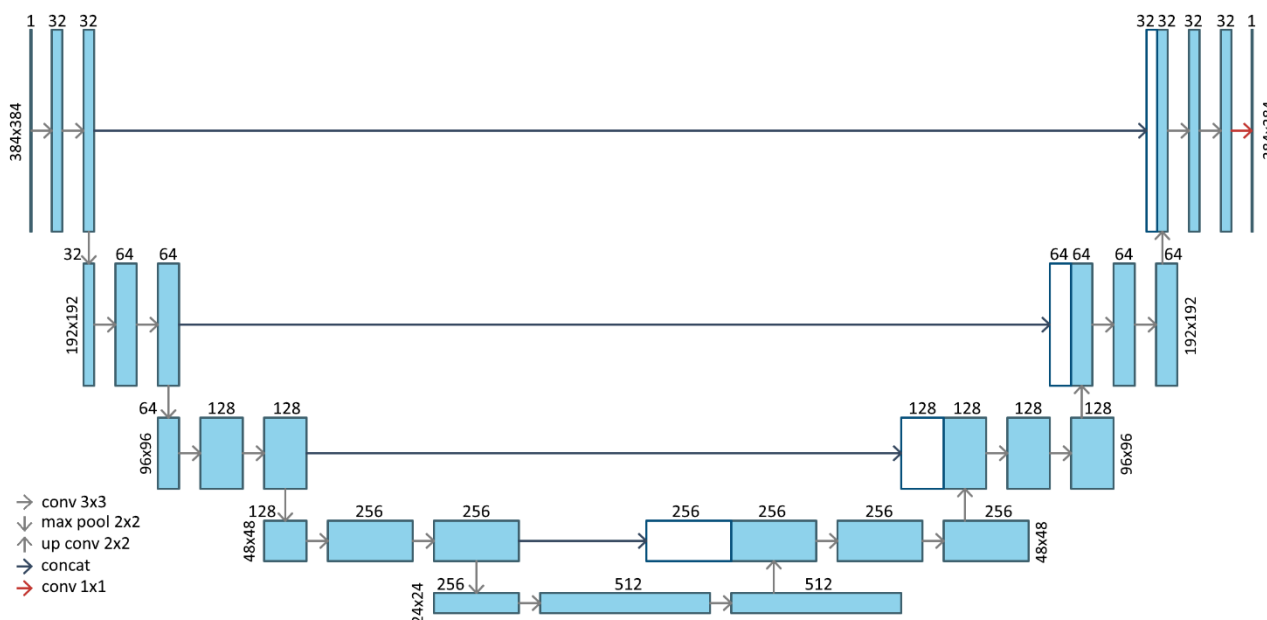


Рисунок 2.5 - Схема конечной сверточной нейронной сети U-Net

Входными значениями являются изображения размерами 384×384 . Первый сверточный слой представляет собой 32 карты сверточного слоя равного размера, получаемых с помощью ядра размером 3×3 . Первый подвыборочный слой представляет собой операцию выборки локальных максимальных значений

с помощью ядра размером 2×2 и шагом, соответствующим размеру ядра. Со второго по пятый сверточный слой происходит увеличение количества карт: 64, 128, 256, 512 соответственно. Со второго по четвёртый подвыборочный слой никаких изменений не происходит.

Далее добавляются слои конкатенации карт признаков из кодирующей части. И вместо сверточных слоёв используются слои обратной свертки, который увеличивает размер изображения в 2 раз по обеим осям. И с шестого по девятый слой обратной свертки происходит уменьшение количества карт: 256, 128, 64, 32 соответственно.

В сверточных слоях и слоях обратной свертки используется функция активации ReLU, расчёт которой производится по формуле 2.1:

$$f(x) = \begin{cases} 0, & x < 0; \\ x, & x \geq 0. \end{cases} \quad (2.1)$$

На последнем уровне свертка используется для формирования выходного сегментированного изображения, равного по размерам входному изображению. Используется сигмоидная функция активации в форме гиперболического тангенса.

Ввиду довольно большого времени обучения конечной сети в некоторых случаях была задействована вторая структура, представленная на рисунке 2.6.

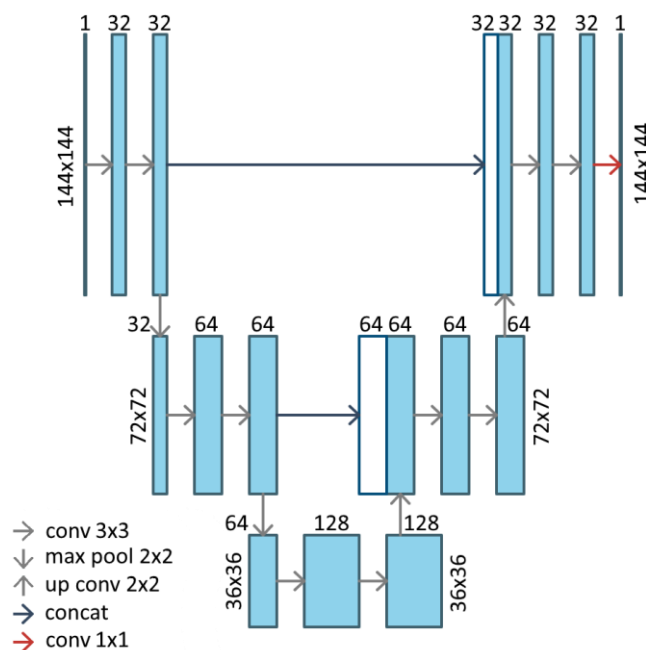


Рисунок 2.6 - Схема проверочной сверточной нейронной сети U-Net

Данная сеть отличается от конечной сети размерами входных изображений, а также количеством слоёв.

2.5 Параметры оценки при обучении сети

Для вычисления ошибки сети по формуле 2.2 рассчитывается коэффициент Дайса (Dice coefficient), который показывает меру сходства изображений.

$$dice\ coefficient = \frac{2 * intersection(y_{true}, y_{pred}) + smooth}{y_{true} + y_{pred} + smooth}, \quad (2.2)$$

где y_{true} - истинное значение пикселей изображений, y_{pred} - предсказанное значение пикселей изображений, $intersection$ - пересечение/умножение значений пикселей изображений, $smooth$ - коэффициент сглаживания.

Чем выше значение коэффициента Дайса, тем большее количество истинных и предсказанных значений пикселей изображений совпадает, соответственно тем лучше детектируются текстовые области.

В качестве дополнительных параметров оценивания работы сети было решено ввести следующие значения, представленные в формулах 2.3-5.

$$Precision = \frac{TP}{TP + FP}, \quad (2.3)$$

где $Precision$ – точность, TP - истинно-положительное решение, FP - ложно-положительное решение.

$$Recall = \frac{TP}{TP + FN}, \quad (2.4)$$

где $Recall$ – полнота, TP - истинно-положительное решение, FN - ложно-отрицательное решение.

$$F = (\beta^2 + 1) \frac{Precision * Recall}{\beta^2 * Precision + Recall}, \quad (2.5)$$

где F – F -мера, β – коэффициент от 0 до ∞ .

2.6 Фильтры для предобработки изображений

2.6.1 Двусторонний фильтр сглаживания

Двусторонний фильтр определяется по формуле 2.6:

$$I^{filtered}(x, y) = \frac{\sum_{j=-N}^N \sum_{i=-N}^N I(i, j) * f_r(I(x, y), I(i, j)) * g_s(x, y, i, j)}{\sum_{j=-N}^N \sum_{i=-N}^N f_r(I(x, y), I(i, j)) * g_s(x, y, i, j)}, \quad (2.6)$$

где $I^{filtered}$ - отфильтрованное изображение; I - входное изображение для фильтрации; (x, y) - координаты текущего пикселя, подлежащего фильтрации; (i, j) - координаты текущего пикселя из маски; N – размер маски; f_r - ядро диапазона для сглаживания разностей интенсивностей (функция Гаусса); g_s - пространственное ядро для сглаживания разностей координат (функция Гаусса).

2.6.2 Фильтр резкости

Фильтрация производится корреляцией исходного изображения с выбранной маской, рассчитываемой по формуле 2.7:

$$I^{filtered}(x, y) = \sum_{j=-N}^N \sum_{i=-N}^N kernel(i, j) * I(x + i, y + j), \quad (2.7)$$

где $I^{filtered}$ - отфильтрованное изображение; I - входное изображение для фильтрации; $kernel$ – маска; N – размер маски.

Маска для фильтра, повышающая резкость, и маска для фильтра резкости, выделяющая края, представлены на рисунке 2.7.

0	-1	0
-1	5	-1
0	-1	0

а

0	-1	0
-1	4	-1
0	-1	0

б

Рисунок 2.7 - Маски для фильтра резкости (а – повышающая резкость, б – выделяющая края)

2.6.3 Фильтр, выделяющий края

Сначала по формуле 2.8 выделяется задний фон путём последовательного применения эрозии и расширения:

$$Bg = (I \ominus kernel) \oplus kernel, \quad (2.8)$$

где I - входное изображение; $kernel$ – маска; Bg – изображение заднего фона.

Затем по формуле 2.9 каждый пиксель исходного изображения делится на соответствующий ему пиксель заднего фона:

$$D(x, y) = \frac{I(x, y)}{Bg(x, y)}, \quad (2.9)$$

где (x, y) – координаты пикселя; D – изображение, полученное при делении.

После чего по формуле 2.10 полученное изображение проходит через пороговую функцию:

$$I^{filtered}(x, y) = \begin{cases} 0, & \text{если } D(x, y) \geq \text{threshold;} \\ 255, & \text{если } D(x, y) < \text{threshold,} \end{cases} \quad (2.10)$$

где $threshold$ – пороговое значения; $I^{filtered}$ - отфильтрованное изображение.

2.7 Преобразование Фурье

В качестве функции для частотного анализа было выбрано преобразование Фурье. Результат применения данной функции состоит в том, что любая функция, периодически воспроизводящая свои значения, может быть представлена в виде суммы синусов и/или косинусов различных частот, умноженных на некоторые коэффициенты [18].

Формула 2.11 представляет собой преобразования Фурье двух переменных:

$$F(u, v) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) e^{-i2\pi(ux+vy)} dx dy, \quad (2.11)$$

Для обратного преобразования используется формула 2.12:

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u, v) e^{i2\pi(ux+vy)} du dv, \quad (2.12)$$

где i – мнимая единица.

2.8 Сглаживающие частотные фильтры

Контурные и другие резкие перепады яркости на изображении (например, связанные с шумом) вносят значительный вклад в высокочастотную часть его

фурье-преобразования [18]. Следовательно, сглаживание достигается в частотной области ослаблением высокочастотных компонент определенного диапазона фурье-образа данного изображения. Фильтрация в частотной области задается равенством, представленным в формуле 2.13:

$$G(u, v) = H(u, v)F(u, v), \quad (2.13)$$

где $H(u, v)$ – функция сглаживания.

2.8.1 Идеальный фильтр низких частот

Идеальный фильтр низких частот задается формулой 2.14:

$$H(u, v) = \begin{cases} 1 & \text{при } D(u, v) < D_0, \\ 0 & \text{при } D(u, v) \geq D_0, \end{cases} \quad (2.14)$$

где $D(u, v)$ – расстояние от точки (u, v) до начала координат, D_0 – заданная дистанция от начала координат.

2.8.2 Фильтр низких частот Гаусса

Фильтр низких частот Гаусса задается формулой 2.15:

$$H(u, v) = e^{-D^2(u, v)/2D_0^2}, \quad (2.15)$$

где $D(u, v)$ – расстояние от точки (u, v) до начала координат, D_0 – заданная частота среза.

2.8.3 Фильтр низких частот Баттерворта

Фильтр низких частот Баттерворта задается формулой 2.16:

$$H(u, v) = \frac{1}{1 + (D(u, v)/D_0)^{2n}}, \quad (2.16)$$

где $D(u, v)$ – расстояние от точки (u, v) до начала координат, D_0 – заданная частота среза, n – заданный порядок.

3 Результаты обучения сети

Для обучения сети на сервисе Google Cloud Platform была создана виртуальная машина с 8 виртуальными центральным процессорами и 104 ГБ памяти. Платформа ЦП не указывается и выбирается системой автоматически.

3.1 Подбор размеров входных изображений

Подбор размеров входных изображений производился на исходной базе изображений KAIST Scene Text Database. Для свёрточных слоёв используется размер ядра 3×3 пикселя, а с помощью подвыборочных слоёв производится уменьшение изображения в 2 раза. Данные параметры учитывались для выбора размеров входных изображений для сравнения. Сравнительное обучение производилось для сети с 10 слоями, поэтому размеры входных изображений должны были без остатка делиться как на 3, так и на 2^4 . Поэтому, для сравнения были выбраны следующие размеры входных изображений: 192×192 пикселя и 384×384 пикселя.

На рисунке 3.1 представлены графики зависимости коэффициента Дайса на обучающей и валидационной выборках от количества эпох. Для изображений размером 192×192 пикселей было проведено обучение в 110 эпох, для изображений размером 384×384 пикселей – 60 эпох.



Рисунок 3.1 – Сравнительное обучение конечной сети для различных размеров входных изображений

Дальнейшее обучение не проводилось ввиду того, что среднее изменение точности для последних 10 эпох не превышало 0,06% и 0,02% соответственно. Следовательно, такие значения несоизмеримы с затраченным временем. В таблице 3.1 представлены конечные параметры обучения.

Таблица 3.1 - Сравнительное обучение конечной сети для различных размеров входных изображений

Размер входного изображения, пиксель	192×192	384×384
Количество эпох	110	60
Время обучения одной эпохи, минута	6,93	10,67
Общее время обучения, час	12,7	10,7
Коэффициент Дайса для обучения	0,5744	0,6025
Коэффициент Дайса для валидации	0,5483	0,5804
Среднее изменение точности последних 10 эпох, %	0,06	0,02

На рисунке 3.2 представлены лучшие маски для тестовых изображений, полученные из сети максимальной точности. Сверху представлен оригинал изображения, в центре – результат обучения размеров 192×192 пикселей, снизу - результат обучения размеров 384×384 пикселей.

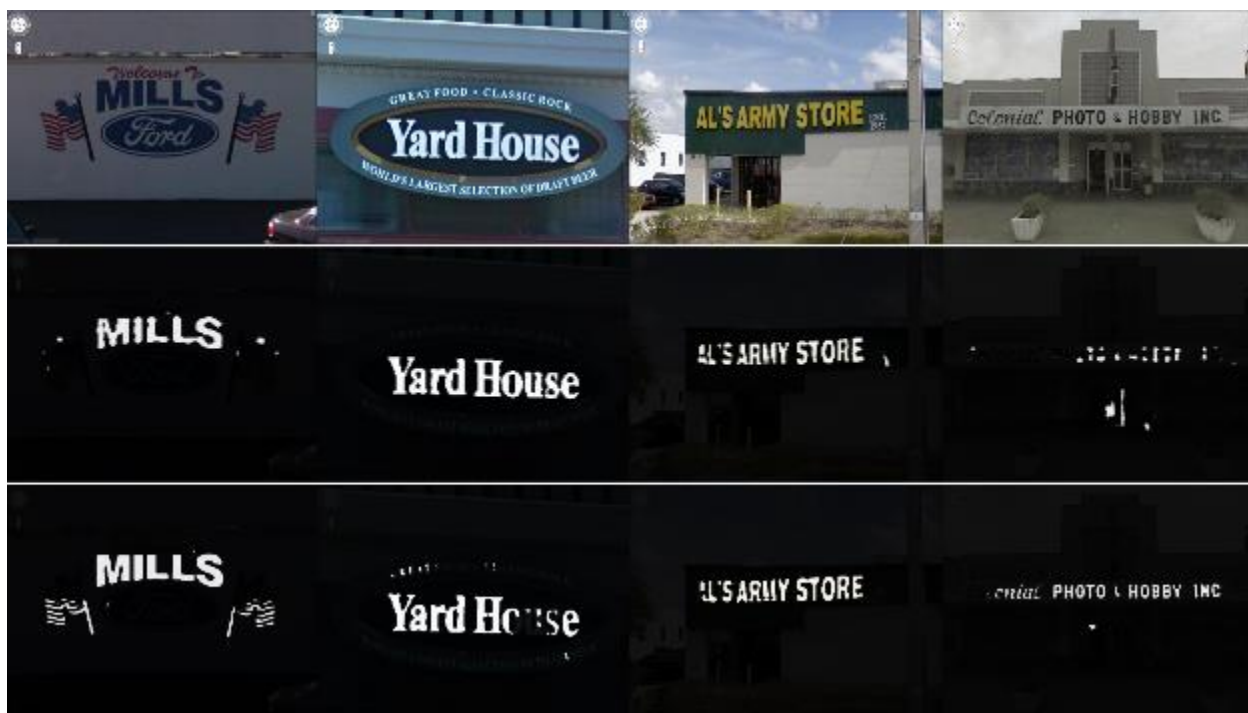


Рисунок 3.2 - Лучшие маски для тестовых изображений, полученные из сети максимальной точности (сверху - оригинал изображения, в центре – результат обучения размеров 192×192 , снизу - результат обучения размеров 384×384)

Видно, что обучение на изображениях размерами 192×192 пикселя лучше справляется с детектированием больших текстовых областей, отсеивая лишние элементы, но не находит маленькие области. В то время как обучение на изображениях размерами 384×384 пикселя ошибочно детектирует элементы фона на изображениях с большими текстовыми областями, но корректно распознаются малые области.

На рисунке 3.3 представлены худшие маски для тестовых изображений, полученные из сети максимальной точности.



Рисунок 3.3 - Худшие маски для тестовых изображений, полученные из сети максимальной точности (сверху - оригинал изображения, в центре – результат обучения размеров 192×192 , снизу - результат обучения размеров 384×384)

Видно, что обучение на изображениях размерами 192×192 пикселя детектирует много нетекстовых областей (например, окна зданий, просветы между деревьями, вывески с рисунками). Обучение на изображениях размерами 384×384 пикселя также детектирует нетекстовые области, однако в меньшем количестве.

Таким образом, было принято решение в дальнейшем подавать на вход сети изображения размером 384×384 пикселя.

3.2 Подбор количества слоёв

Следующим шагом формирования структуры конечной свёрточной сети был выбор количества слоёв. Было проведено сравнительное обучение для сети с 6ю, 8ю и 10ю слоями. Большее количество слоёв не бралось в рассмотрение ввиду увеличения времени обучения. Во всех представленных случаях на вход сети подавались изображения, приведённые к размерам 384×384 пикселя. Обучение производилось в течении 10 эпох. На рисунке 3.4 представлены графики зависимости коэффициента Дайса обучения и валидации от количества эпох.

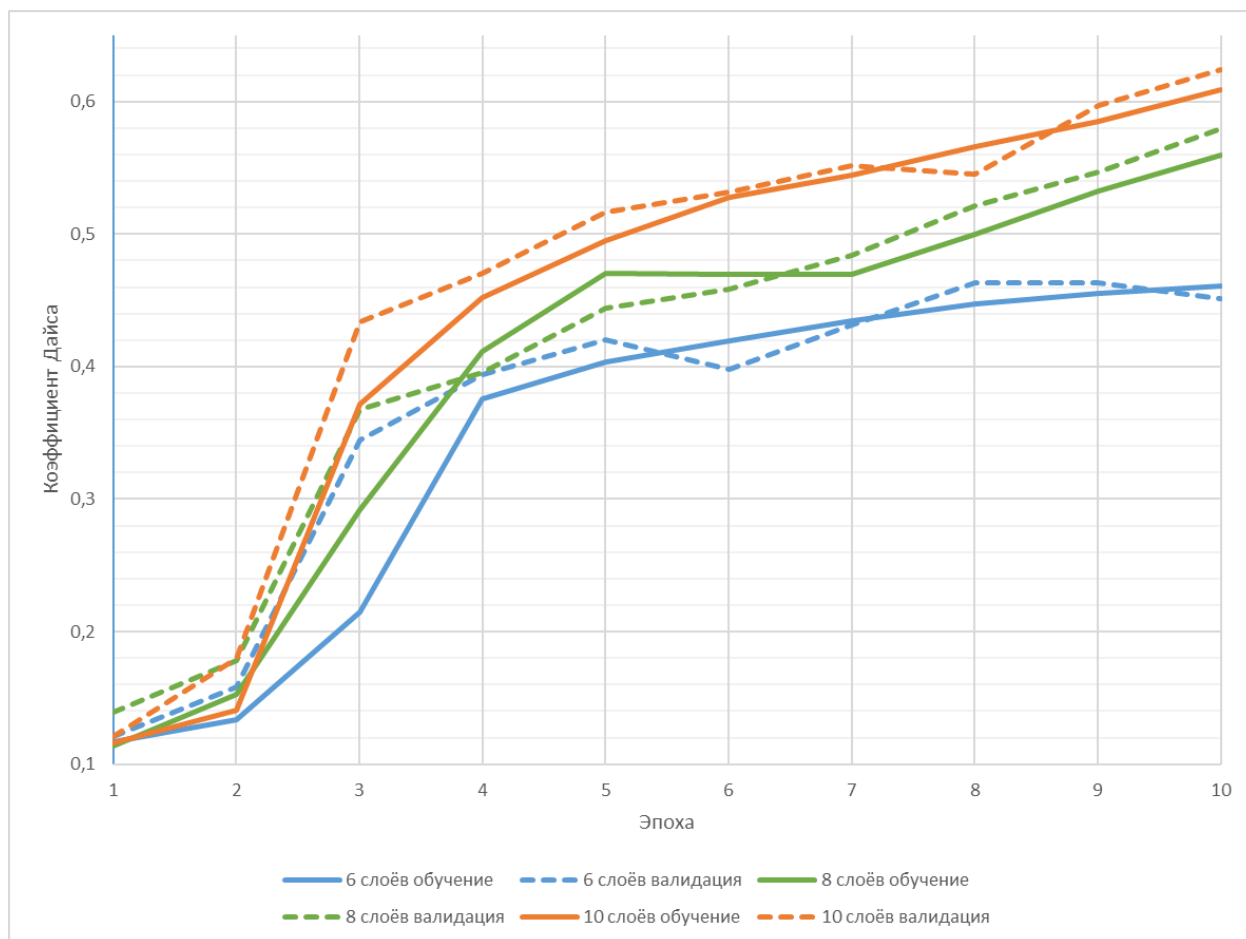


Рисунок 3.4 - Графики зависимости коэффициента Дайса обучения и валидации от количества эпох для 6, 8 и 10 слоёв

На момент последней эпохи коэффициента Дайса для обучения с 6ю, 8ю и 10ю слоями достигает 46,1%, 55,9% и 60,9% соответственно. Видно, что наилучшие результаты даёт сеть с использованием 10 слоёв. Дальнейшее проверочное обучение не требуется, так как в течение последних 5 эпох обучения значения коэффициента Дайса для сети с 10ю слоями в среднем превышают значения для сети с 6ю и 8ю слоями на 27,7% и 12,1% соответственно.

3.3 Подбор типов свёрточных слоёв

Было проведено сравнение для конфигураций сети с использованием слоёв Max Pooling и без них. Во втором случае свёртка слоёв происходила за счёт слоёв Convolution2D. Обучение производилось в течение 10 эпох. На рисунке 3.5 представлены графики зависимости коэффициента Дайса обучения и валидации от количества эпох.

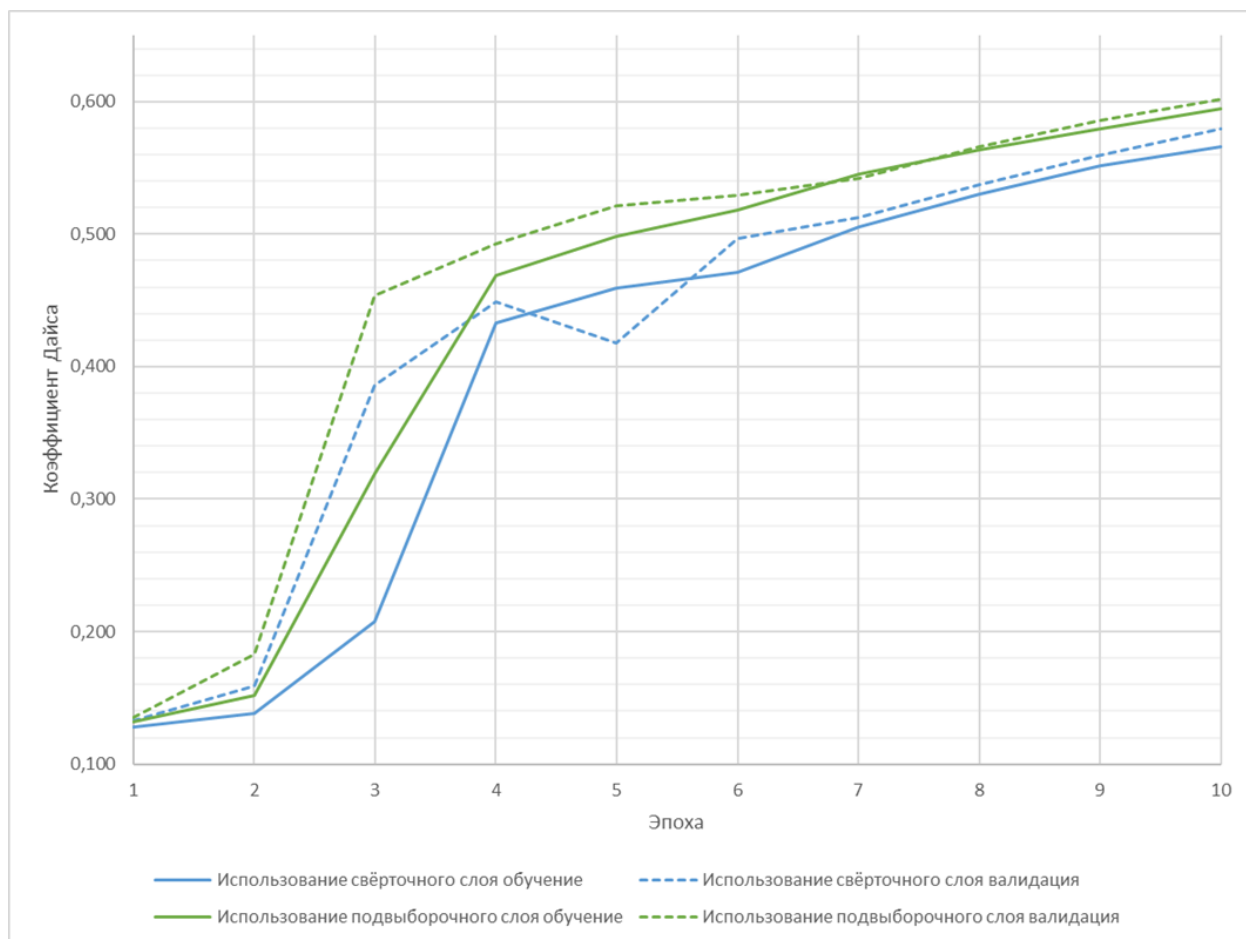


Рисунок 3.5 - Графики зависимости коэффициента Дайса обучения и валидации от количества эпох при использовании слоёв Max Pooling и без данных слоёв

На момент последней эпохи коэффициент Дайса обучения при использовании слоёв Max Pooling и без данных слоёв достигает 59,5% и 56,6% соответственно. Видно, что лучшие результаты обучения достигаются при использовании слоёв Max Pooling.

3.4 Предобработка входных изображений

Было решено использовать предобработку перед подачей изображений в качестве входных данных для сверточной нейронной сети. Используемые фильтры описаны в пункте 2.6. На рисунке 3.6 представлены варианты использования фильтров для случайно выбранного изображения из базы изображений.



Рисунок 3.6 - Предобработка изображения: а – без фильтров, б – двусторонний фильтр сглаживания с коэффициентом 50, в – двусторонний фильтр сглаживания с коэффициентом 50 и фильтр резкости, г - двусторонний фильтр сглаживания с коэффициентом 150, д - двусторонний фильтр сглаживания с коэффициентом 50 и фильтр резкости, выделяющей края, е - двусторонний фильтр сглаживания с коэффициентом 50 и фильтр выделения краёв

Для сравнения предобработки входных изображений было произведено обучение на проверочной сети, описанной в пункте 2.4, на исходной базе изображений длиной в 20 эпох. На рисунках 3.7 и 3.8 представлены графики зависимости коэффициента Дайса от количества эпох соответственно.

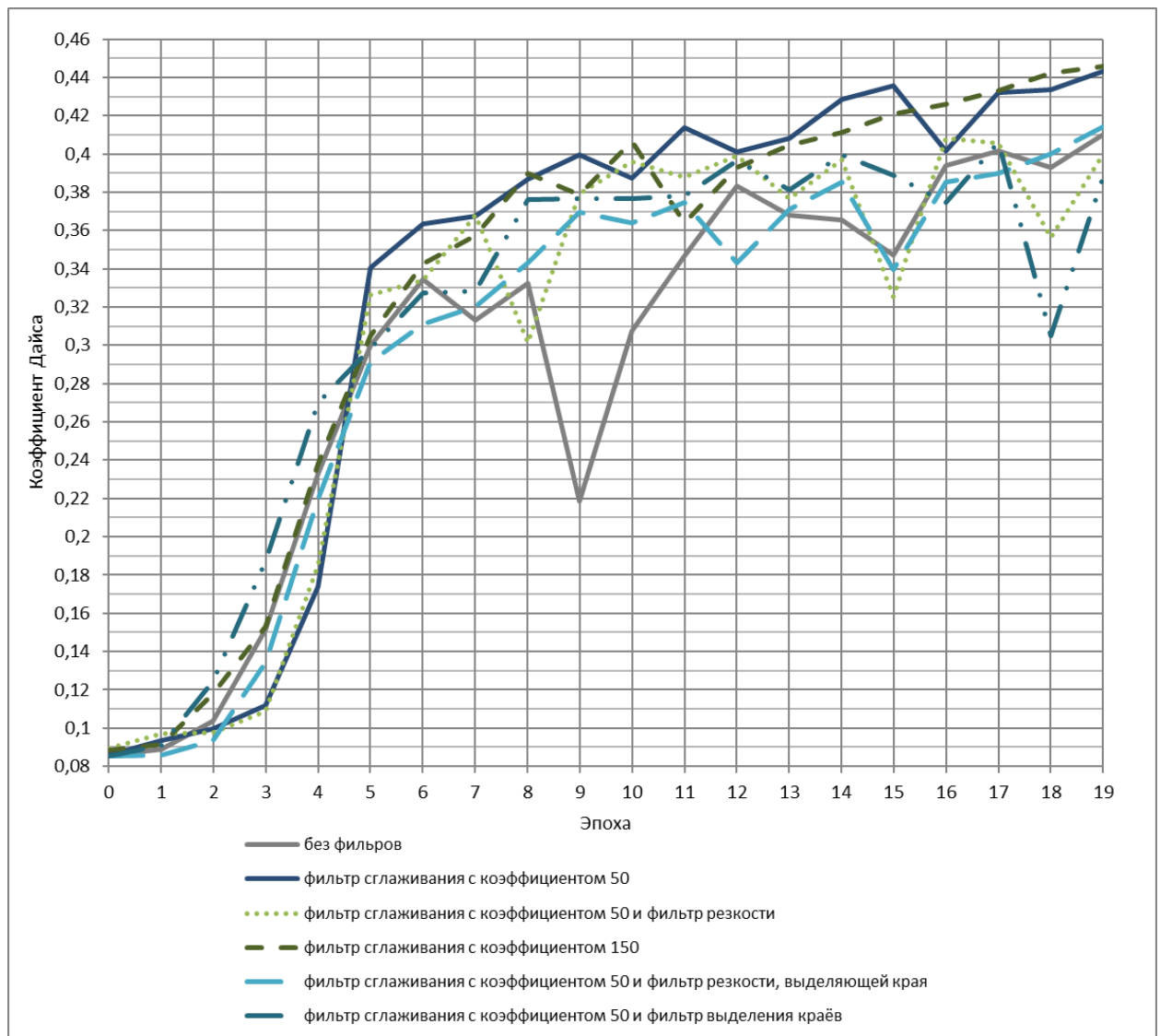


Рисунок 3.7 - Графики зависимости коэффициента Дайса обучения от количества эпох для различных фильтров

Видно, что для обучающей выборки наиболее точными и стабильными являются двусторонний фильтр сглаживания с коэффициентом 50 и с коэффициентом 150.

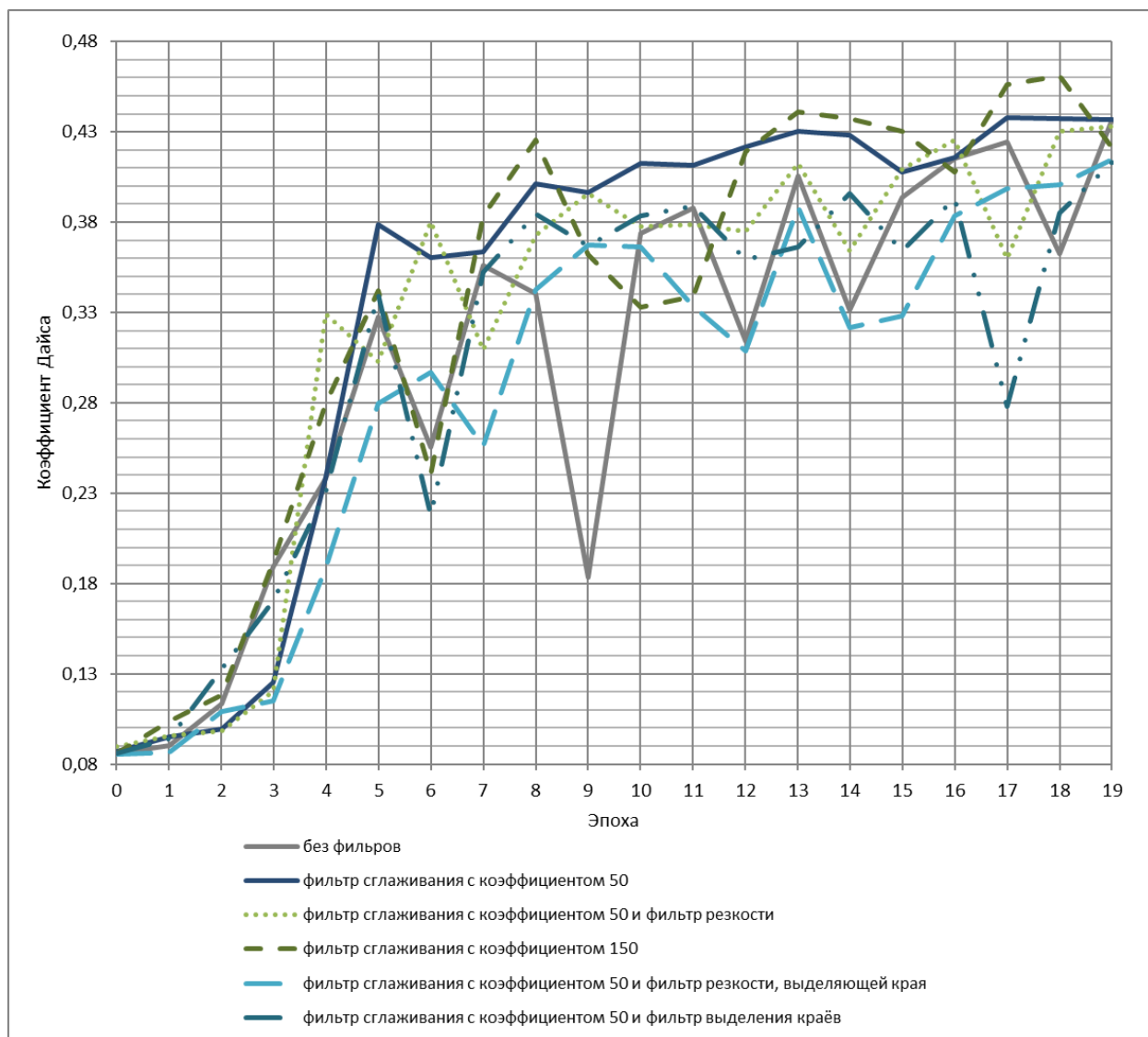


Рисунок 3.8 - Графики зависимости коэффициента Дайса валидации от количества эпох для различных фильтров

Для валидационной выборки большинство фильтров достаточно непостоянны, самым стабильным является двусторонний фильтр сглаживания с коэффициентом 50.

Таким образом было решено использовать двусторонний фильтр сглаживания с коэффициентом 50 для обучения конечной сети и сравнения его результатов с обучением без фильтров. Для этого была использована база изображений, увеличенная с помощью вращений. На рисунке 3.9 представлены графики зависимости коэффициента Дайса от количества эпох при использовании фильтра сглаживания и без него.

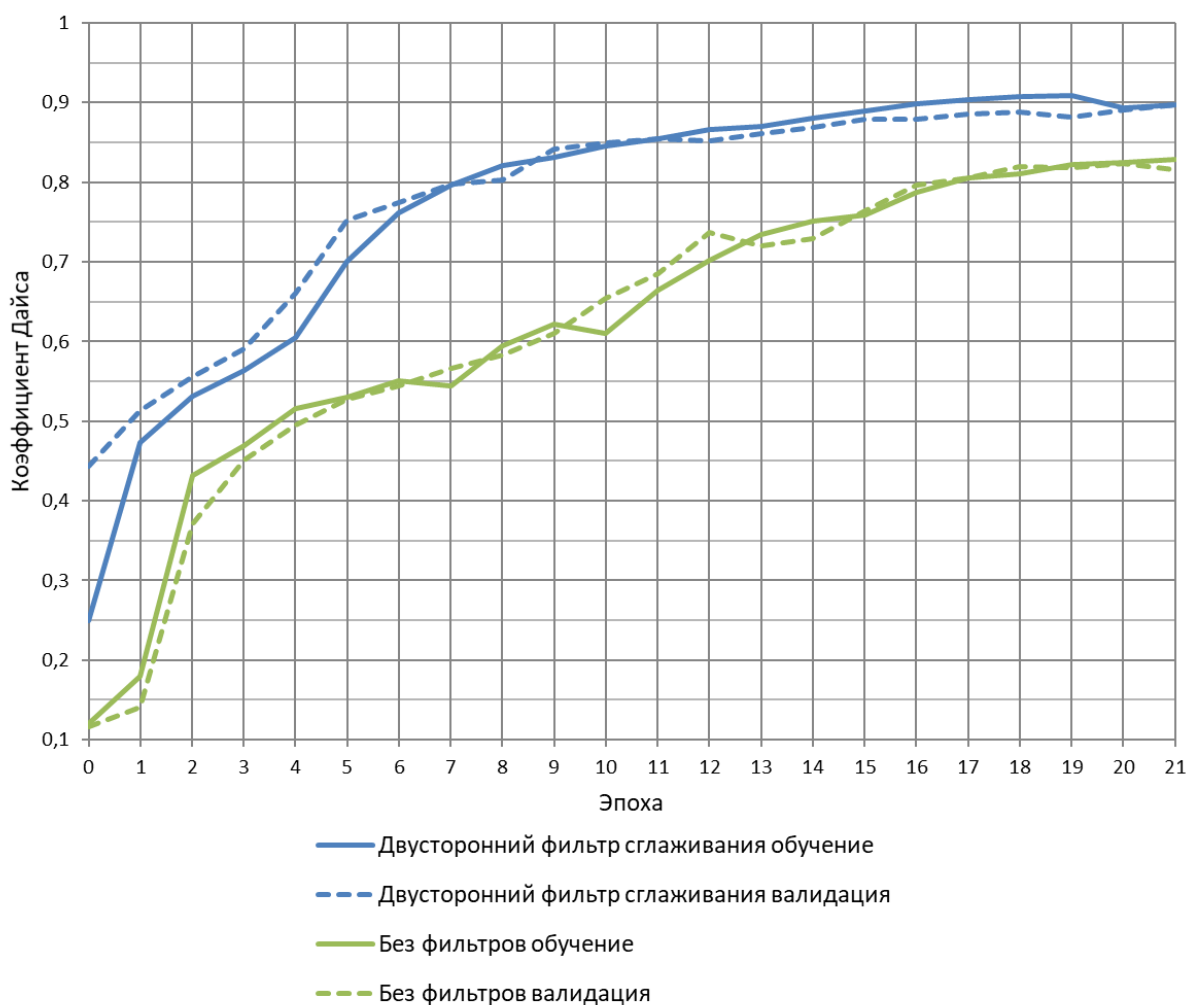


Рисунок 3.9 - Графики зависимости коэффициента Дайса обучения и валидации от количества эпох при использовании фильтра сглаживания и без него

На момент последней эпохи коэффициенты Дайса для обучения, валидации и тестирования представлены в таблице 3.2.

Таблица 3.2 - Сравнительное обучение конечной сети для с использованием двустороннего фильтра сглаживания для входных изображений и без него

Параметр		С фильтром сглаживания	Без фильтра
Коэффициент Дайса	Обучение	0,897	0,829
	Валидация	0,897	0,816
	Тестирование	0,894	-

Фильтрация размытием позволила убрать шумы на изображениях, что облегчило сети выделение контуров текстовых областей.

На рисунках 3.10 и 3.11 представлены лучшие и худшие маски для тестовых изображений, полученные из сети максимальной точности обучения с фильтрацией.



Рисунок 3.10 - Лучшие маски для тестовых изображений, полученные из сети максимальной точности (сверху - оригинал изображения, снизу – результат обучения с фильтрацией)

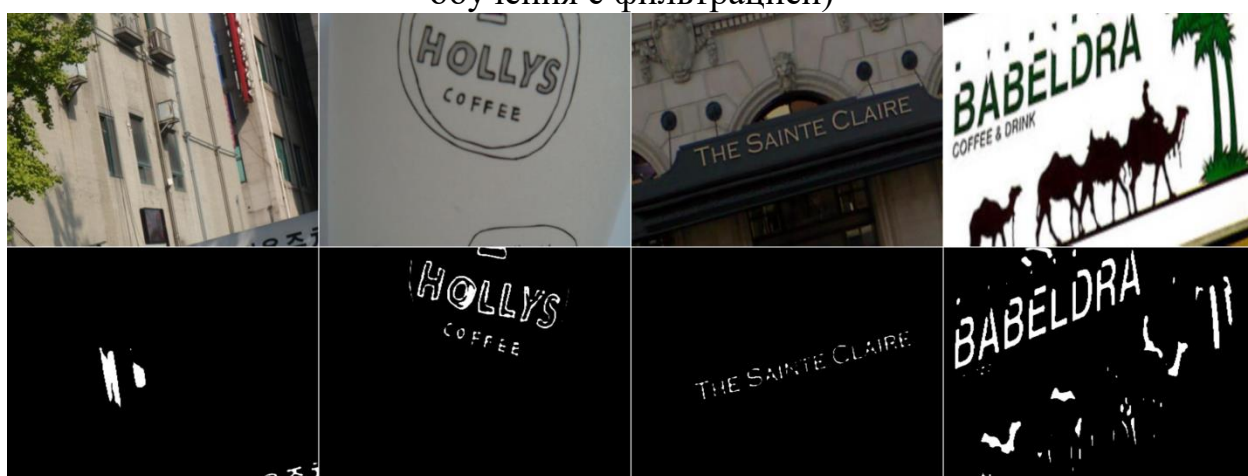


Рисунок 3.11 - Худшие маски для тестовых изображений, полученные из сети максимальной точности (сверху - оригинал изображения, снизу – результат обучения с фильтрацией)

Несмотря на то, что на большей части тестовых изображений текстовые области детектируются правильно, в некоторых случаях сеть выделяет такие области как, окна зданий, просветы между деревьями, вывески с рисунками. Некоторые текстовые надписи с тонким шрифтом так же плохо определяются за счёт сжатия изображений. Неоднородные текстовые надписи (градиент, контурные шрифты) также представляют собой проблему.

2.5 Частотный анализ с использованием преобразования Фурье

В качестве альтернативного варианта предобработки входных изображений был рассмотрен частотный анализ с преобразованием Фурье.

2.5.1 Подача на вход сети фрагментов изображения после преобразования Фурье

Для применения преобразования Фурье была взята исходная база изображений, и произведено её разрезание на фрагменты размерами 100×100 пикселей. Полученный набор изображений был разделён на фрагменты с текстом, представленные на рисунке 3.12, и фрагменты без текста, представленные на рисунке 3.13. Количество используемых фрагментов обеих групп было приведено к равному значению для балансировки выборок.



Рисунок 3.12 - Фрагменты изображений с текстом



Рисунок 3.13 - Фрагменты изображений без текста

Для подачи на вход обучаемой сети ко всем изображениям из полученной базы было применено преобразование Фурье. Для сравнения обучений также были сформированы выборки с центрированием полученных преобразований. Эталонным выходным значением для каждого входного изображения являются значения 0 или 1, в зависимости от отсутствия или наличия текстовых областей соответственно.

Пример результатов обучения сети длиной в 10 эпох при подаче на вход сети фрагментов изображений с преобразованием Фурье представлен на рисунке 3.14.

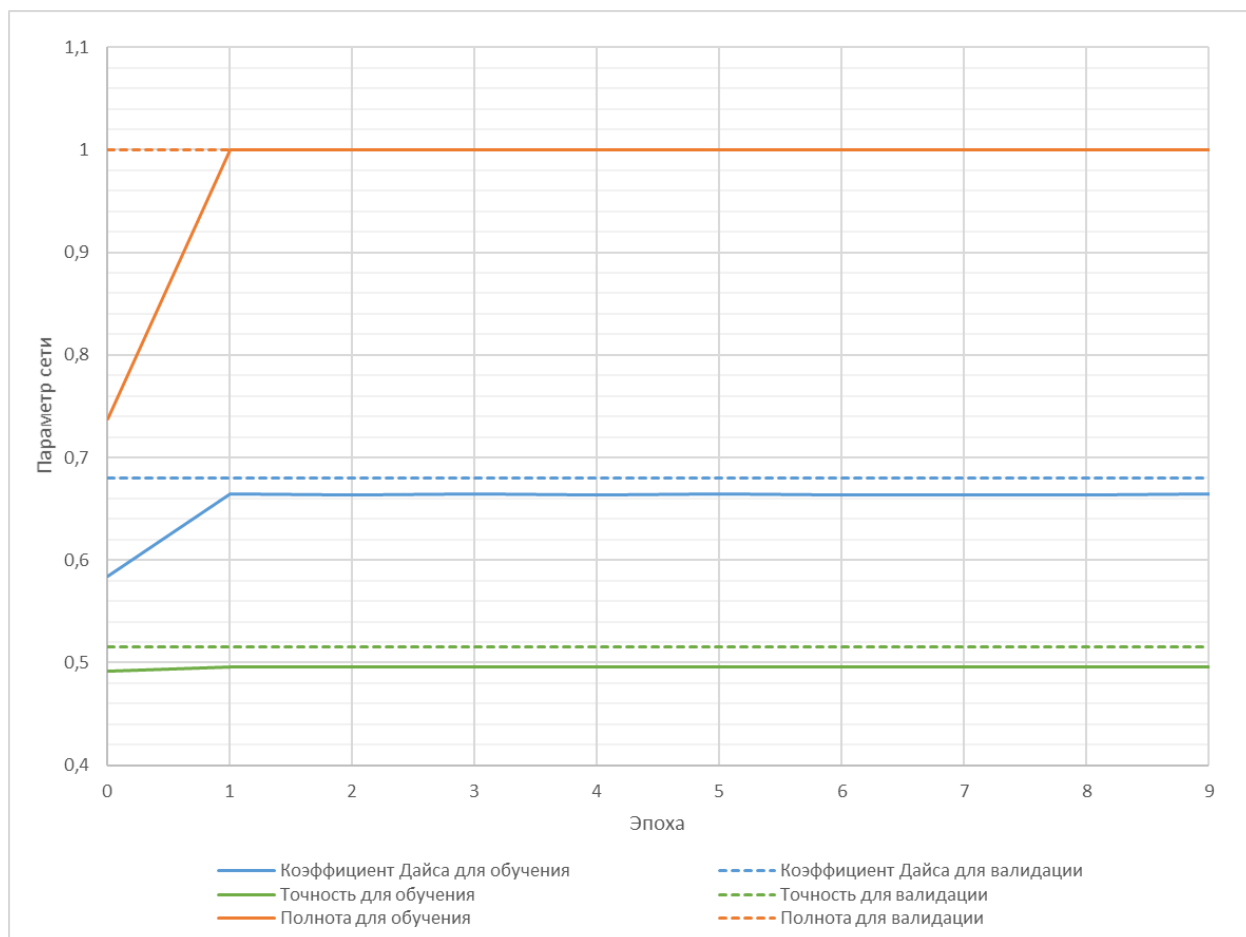


Рисунок 3.14 - Результат обучения сети при подаче на вход сети фрагментов изображений с преобразованием Фурье

Видно, что полнота как для обучения, так и для валидации приняла значение, равное 1, что означает предсказания всех выходных значений сети как 1, то есть содержащих текстовые области.

Повторные попытки обучения, а также попытки обучения с центрированными версиями преобразования Фурье показывали аналогичные результаты. Также были показаны результаты, при которых полнота принимала значение 0, что означает предсказания всех выходных значений сети как 0, то есть не содержащих текстовые области.

3.5.2 Добавление слоёв с преобразованием Фурье

Для реализации добавления слоёв с преобразованием Фурье в начало сети использовались функции библиотеки tensorflow. На рисунке 3.15 показаны результаты проверочного обучения нейронной сети для 2 попыток применения

слоёв с преобразованием Фурье. Для сравнения добавлены результаты обучения без данных слоёв.

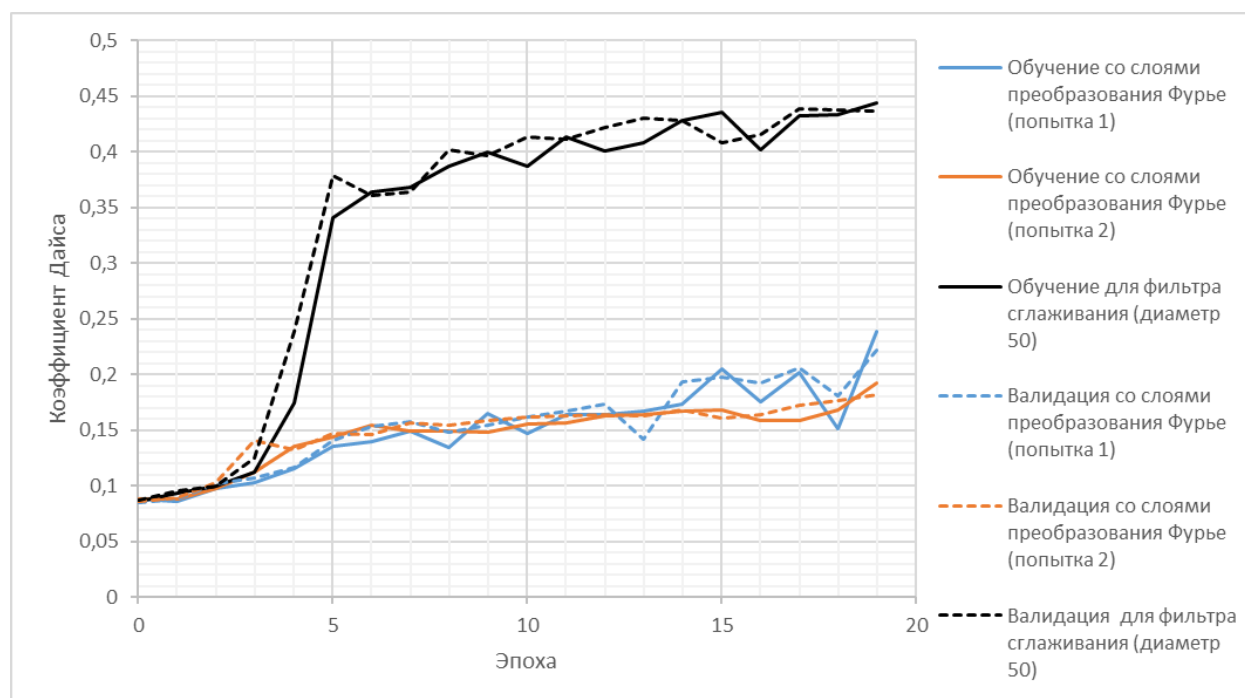


Рисунок 3.15 - Результаты проверочного обучения нейронной сети для 2 попыток применения слоёв с преобразованием Фурье

Видно, что добавление данных слоёв значительно снизило значения коэффициента Дайса. То есть, обучение сети не производится.

2.5.3 Сглаживающие частотные фильтры с преобразованием Фурье

Для использования преобразования Фурье как части сглаживающих фильтров были выбраны фильтры, перечисленные в пункте 2.8. Их параметры, представленные в таблице 3.3.

Таблица 3.3 - Фильтры размытия и их параметры

Название фильтра низких частот	Значения параметра D_0	Значения параметра n
Идеальный фильтр	80, 120, 160, 200	-
Фильтр Гаусса	120, 160, 200	-
Фильтр Баттерворта	120, 160, 200	2, 5, 8

На рисунке 3.16 представлены примеры фильтров для значения параметра D_0 равного 120.

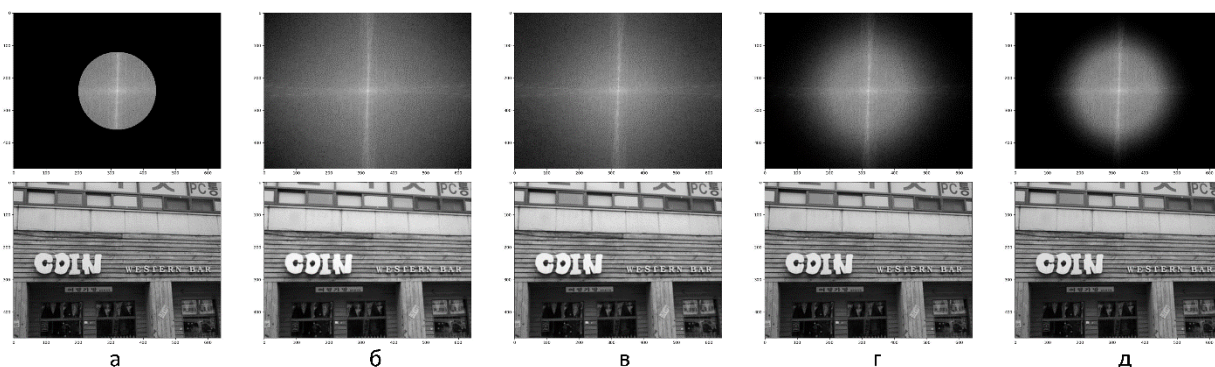


Рисунок 3.16 - Примеры фильтров и их применения для параметра $D_0=120$ (а - идеальный фильтр низких частот, б - фильтр низких частот Гаусса, в - фильтр низких частот Баттерворта низких частот ($n=2$), г - фильтр низких частот Баттерворта ($n=5$), д - фильтр низких частот Баттерворта ($n=8$))

Результаты обучения сети для каждого фильтра отдельно представлены в приложении Б. В результате были выбраны лучшие значения параметров и проведено их сравнение с фильтром сглаживания. На рисунке 3.17 показано сравнение результатов проверочного обучения нейронной сети для лучших фильтров с преобразованием Фурье и фильтра сглаживания.

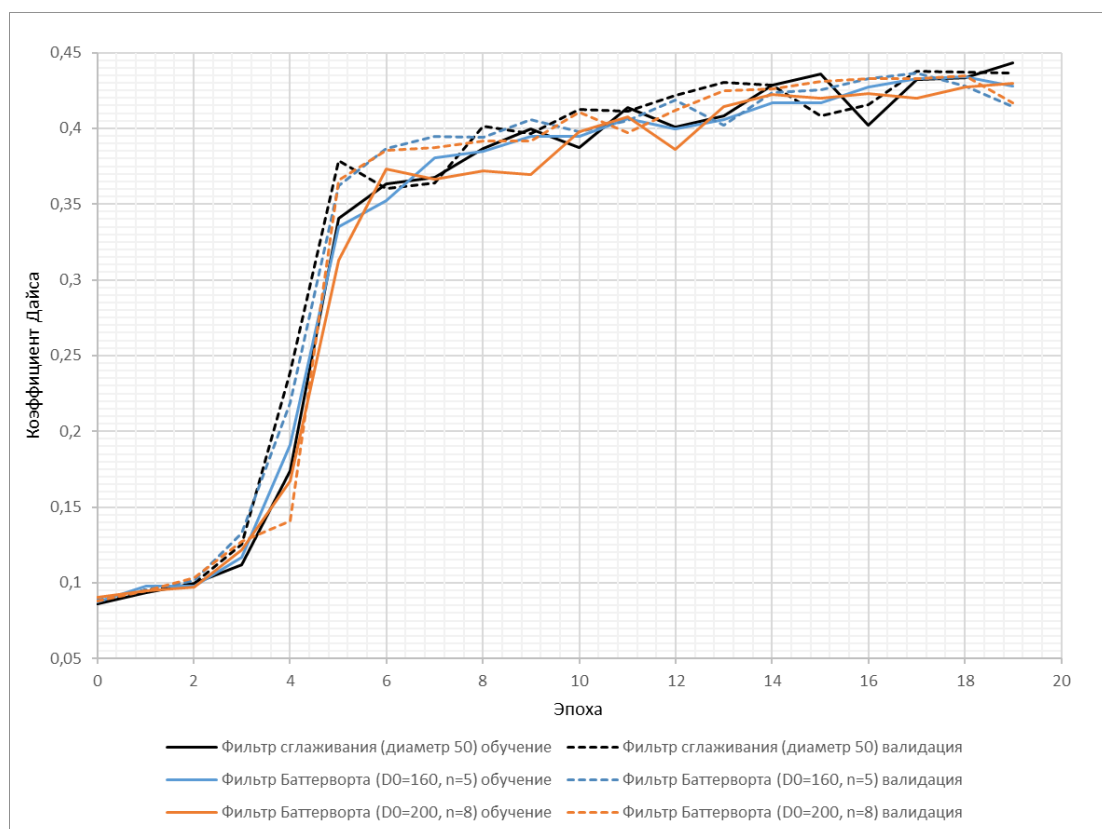


Рисунок 3.17 - Сравнение результатов проверочного обучения сети для лучших фильтров с преобразованием Фурье и фильтра сглаживания

Видно, что коэффициенты Дайса даже лучших фильтров с преобразованием Фурье (фильтр Баттерворта с $D_0=160$ и $n=5$, фильтр Баттерворта с $D_0=200$ и $n=8$) не превышают значения коэффициентов для фильтра сглаживания.

2.6 Разбиение и сжатие изображений перед подачей на вход сети

Так как исходный размер изображений равен 640×480 пикселей, а на вход сети подаются изображения размером 384×384 пикселя, то было решено попробовать разбиение изображений на 4 части вместо изменения размеров. На рисунке 3.18 показан пример сжатия и разбиения на 4 части одного и того же изображения.



Рисунок 3.18 - Пример предварительной обработки изображения (слева – изображение, сжатое до размеров 384×384 пикселя, справа – части изображения размером 384×384 пикселя)

Было проведено проверочное обучение длиною в 10 эпох для 3 вариантов подачи входных изображений: сжатие, разделение, комбинация сжатия и разделение. На рисунке 3.19 представлены графики коэффициента Дайса для каждого варианта.

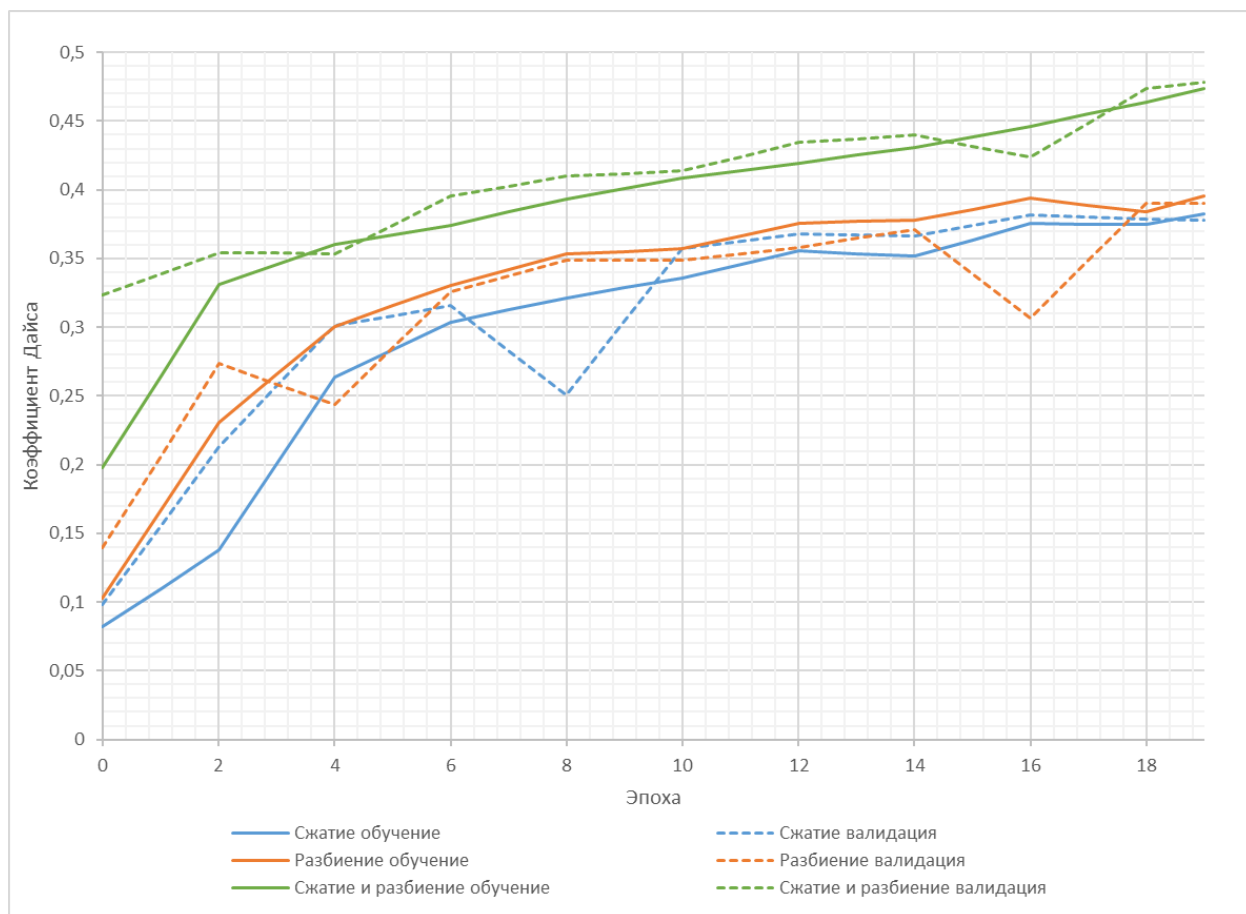


Рисунок 3.19 - Сравнение результатов проверочного обучения сети для 3 вариантов подачи входных изображений: сжатие, разбиения, комбинация сжатия и разбиения

Видно, что наилучший результат показывает комбинация сжатия и разбиения, поэтому для данного варианта было проведено обучение конечной структуры сети на полной базе изображений длиной в 5 эпох. На рисунке 3.20 представлены графики зависимости коэффициента Дайса от количества эпох при таком формате входных данных.

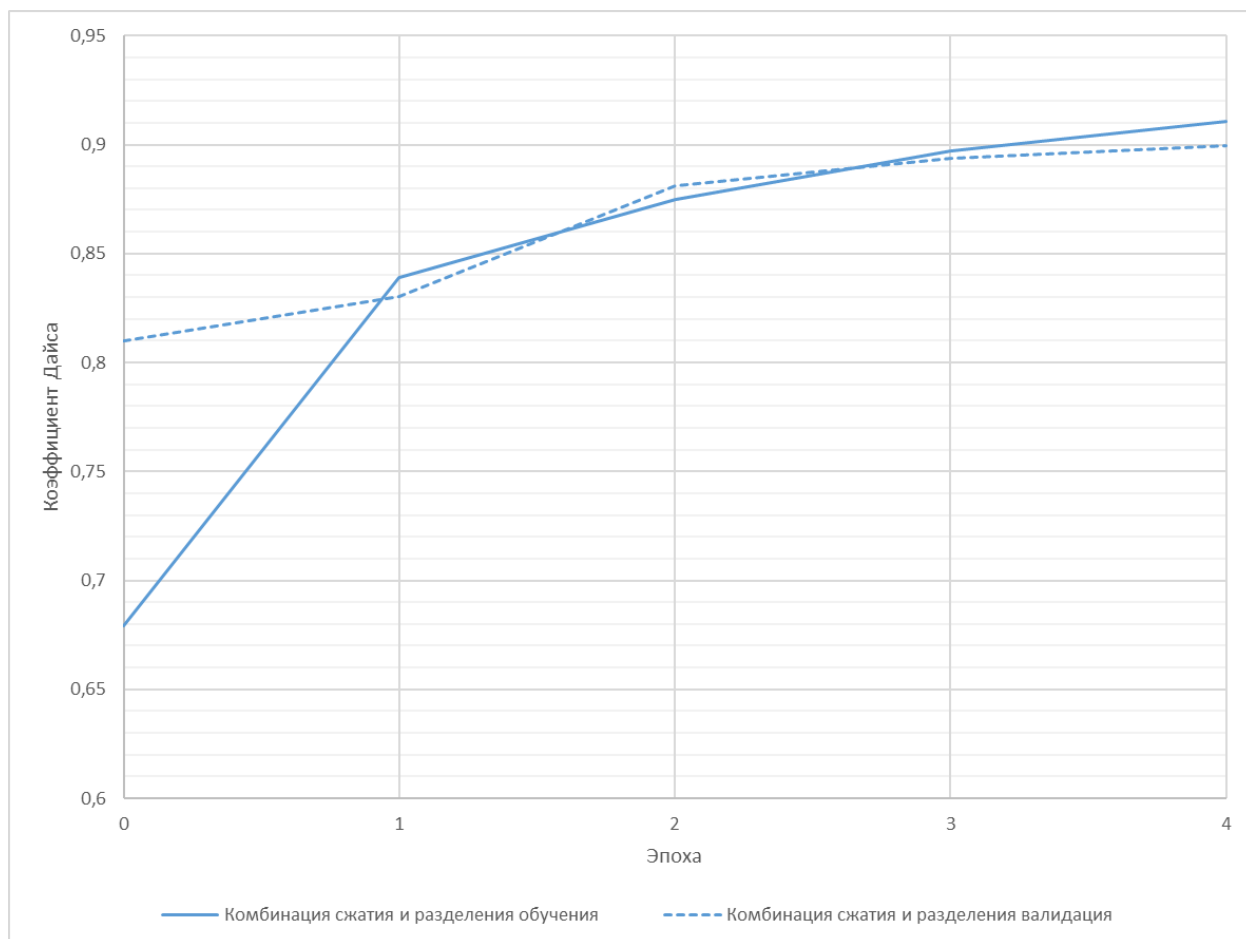


Рисунок 3.20 - Графики зависимости коэффициента Дайса обучения и валидации от количества эпох для комбинация сжатия и разбиения входных изображений

В таблице 3.4 представлены параметры оценки обучения, валидации и тестирования. Сравнение производится обучением на конечной сети при подаче только сжатых изображений.

Таблица 3.4 – Сравнение обучения конечной сети со сжатием входных изображений и с комбинацией сжатия и разбиения

Параметр	Тип	Конечная сеть, база изображений с поворотами и сжатием		Конечная сеть, база изображений с поворотами, сжатием и разбиением
Количество эпох обучения		22	5	5
Полное время обучения, час		59	13,5	60
Коэффициент Дайса	Обучение	0,897	0,605	0,911
	Валидация	0,897	0,660	0,900
	Тестирование	0,894	-	0,880
Точность	Обучение	0,948	0,584	0,915
	Валидация	0,945	0,813	0,917
	Тестирование	0,946	-	0,902
Полнота	Обучение	0,807	0,653	0,902

F-мера ($\beta=1$)	Валидация	0,823	0,681	0,882
	Тестирование	0,822	-	0,864
	Обучение	0,8718	0,631	0,908
	Валидация	0,8798	0,662	0,900
	Тестирование	0,918	-	0,883

Результаты показывают, что за одинаковое количество эпох комбинация сжатия и разделения существенно увеличивает значения параметров оценки сети как для обучающей, так и для валидационной выборки. Сравнение с результатами предсказаний для тестовой выборки не может быть произведено ввиду проведения тестирования после полного завершения обучения сети.

При сравнении с результатами обучения при равном времени обучения сетей видно, что комбинация сжатия и разбиения позволяет получить лучшие результаты для коэффициента Дайса, полноты и F-меры. Точность уступает на 3,3%, 2,9% и 4,4% для обучающей, валидационной и тестовой выборок соответственно.

Высокие значения полноты указывают на то, что большая часть текстовых областей действительно определяется алгоритмом; в то время как высокие значения точности указывают на то, что большая часть выделенных областей действительно являются текстовыми. Рассмотрение точности и полноты для результатов обучения при равном времени показывает, что для обучения на сжатых изображениях данные значения различаются в среднем на 12,9%, а для обучения на комбинации сжатых и разделенных – на 2,9%. Это означает, что при обучении на сжатых изображениях сеть не способна выделить все текстовые области.

На рисунках 3.21 и 3.22 представлены лучшие и худшие варианты предсказаний соответственно.



Рисунок 3.21 – Лучшие маски для тестовых изображений, полученные из сети максимальной точности



Рисунок 3.22 – Худшие маски для тестовых изображений, полученные из сети максимальной точности

На большей части тестовых изображений текстовые области детектируются правильно. Сеть хорошо справляется с различными типами и размерами шрифтов, а также с поворотами. Сложными случаями для определения являются большие надписи: выделяются дополнительные области вокруг или, наоборот, опускаются области внутри символов.

3.7 Сравнение с другими работами

Было проведено сравнение результатов тестирования текущего алгоритма с результатами, полученными другими авторами при использовании базы изображений KAIST Scene Text Database. Ввиду отсутствия результатов работ других авторов, использующих нейросетевые подходы, были выбраны работы с другими методами.

Список методов, представленных в данных работах, указан ниже:

1. Метод, основанный на квадратичной дискриминантной функции [19];
2. Метод наименьшего разреза графа [19];
3. Комбинация метода квадратичной дискриминантной функции и наименьшего разреза графа [19];
4. Метод определения пороговых значений Otsu [19];
5. Метод определения пороговых значений Niblack [19];
6. Метод на основе кластеризации К-средних [19];
7. Метод максимально стабильных экстремальных областей [20];
8. Модифицированный метод на основе кластеризации К-средних [20];
9. Метод детектирования особых областей (Blob) [20];
10. Метод, основанный на Марковской модели случайного поля [20];
11. Комбинация метода максимально стабильных экстремальных областей и алгоритма дендрограмм одинарной кластеризации [21].

Тестирование работ 1-6 производилось только на английской части базы изображений, в то время как работы 7-11 рассматривают полный набор изображений. Дальнейшее обращение к работам производится по названиям их методов.

В таблице 3.5 представлены параметры оценки тестирования для всех перечисленных методов, а также для метода, представленного в данной работе. В качестве параметров сравнения были использованы точность, полнота и F-мера.

Таблица 3.5 – Сравнение результатов тестирования методов на базе изображений KAIST Scene Text Database

База изображений	Метод	Точность	Полнота	F-мера
Английская часть	Метод, основанный на квадратичной дискриминантной функции	0,789	0,925	0,851
	Метод наименьшего разреза графа	0,786	0,918	0,847
	Комбинация метода квадратичной дискриминантной функции и наименьшего разреза графа	0,828	0,893	0,860

	Метод определения пороговых значений Otsu	0,747	0,905	0,818
	Метод определения порога Niblack	0,689	0,909	0,784
	Метод на основе кластеризации K-средних	0,763	0,914	0,821
Полная	Метод максимально стабильных экстремальных областей	0,924	0,341	0,480
	Модифицированный метод на основе кластеризации K-средних	0,690	0,600	0,640
	Метод детектирования особых областей (Blob)	0,879	0,489	0,593
	Метод, основанный на Марковской модели случайного поля	0,697	0,291	0,376
	Комбинация метода максимально стабильных экстремальных областей и алгоритма дендрограмм одинарной кластеризации	0,670	0,890	0,760
	Текущий метод	0,902	0,864	0,883

Как видно из таблицы выше максимальным значением точности обладает работа с использованием метода максимально стабильных экстремальных областей. Полученное значение превышает точность метода, представленного в данной работе, на 2,5%. Однако полнота и F-мера данного метода показывают очень низкие результаты, следовательно, большое количество текстовых областей не обнаруживается. Все остальные методы показывают точность тестирования ниже, чем текущий метод.

Все методы, тестирование которых проходит на английской части базы изображений, обладают более высокими значениями полноты. При этом разница с текущим методом не превышает 7,0% для наилучшего результата. Это может быть обусловлено более простым набором данных для тестирования. Тем не менее, эти же методы показывают более низкие варианты для точности, что означает наличие большего числа ложных срабатываний на нетекстовых областях.

Среди методов, тестирование которых проходит на полной базе изображений, лишь метод комбинация использования максимально стабильных экстремальных областей и алгоритма дендрограмм одинарной кластеризации показывает значение выше текущего. Но в то же время, точность данного метода является самой низкой из всех представленных значений.

Текущий метод показывает наибольшее значение F-меры. В среднем методы, производящие тестирование на английской части базы изображений, показывают значения на 6,0% меньше; для полной базы количество процентов возрастает до 35,5%.

Таким образом, метод, представленный в данной работе, не обладает наилучшими значениями точности и полноты при их раздельном рассмотрении, но при этом превосходит все остальные по значению F-меры. Это означает, что ни один метод из других работ не превосходит текущий по всем параметрам.

Высокие значения сразу всех параметров показывают, что обучение происходило более сбалансировано, то есть большая часть текстовых областей действительно определяется алгоритмом, и большая часть выделенных областей действительно являются текстовыми. Таким образом, предложенный алгоритм превосходит алгоритмы других авторов по F-мере, что говорит о лучшей его сбалансированности между верными и ложными срабатываниями, и является сопоставимым с лучшими результатами алгоритмов по точности и полноте.

4 Финансовый менеджмент, ресурсоэффективность и ресурсосбережение

4.1 Предпроектный анализ

4.1.1 Анализ конкурентных технических решений

Результатом разработки является нейросетевой алгоритм распознавания надписей на изображениях реальных сцен. Большинство конкурентных решений представлено в виде статей с описанием применяемых методов без указания программного кода и возможности установки описанного программного обеспечения, либо в виде программного кода с нейронной сетью, которую необходимо дополнительно обучать.

В данной работе используется структура сверточной нейронной сети U-Net, обучение которой производится на базе изображений KAIST Scene Text Database. Ввиду отсутствия прямых конкурентных алгоритмов, применяющих как используемую структуру сети, так и используемую базу изображений, сравнение производится с решениями, использующими только указанную структуру сети. Были выбраны алгоритмы, описанные в следующих статьях:

1. «U-Net-bin: hacking the document image binarization contest», в котором производится детектирование рукописного текста [17];
2. «Page Segmentation using a Convolutional Neural Network with Trainable Co-occurrence Features», в котором используется изменённая структура U-Net, а обучение производится на базе изображений с отсканированными страницами текста [22];
3. «Fully Convolutional Neural Networks for Page Segmentation of Historical Document Images», в котором также используется изменённая структура U-Net и база изображений с отсканированными страницами текста [23].

Были выбраны следующие критерии конкурентоспособности:

- Доступность программного кода (наличие в открытом доступе описанного в статье алгоритма в виде программного кода или структуры),

- Точность обучения алгоритма (значения конечных показателей точности обученного сети на выборках данных),
- Потребность в вычислительных ресурсах (уровень требований к вычислительной машине, на которой производится обучение сети),
- Возможность переобучения (возможность обучения сети на других данных или дообучения текущего состояния новыми данными),
- Универсальность относительно входных данных (возможность подачи на вход обученной сети различных входных данных, например, изображения реальных сцен, отсканированные страницы документов или рукописных текстов).

Был оценен вес каждого критерия, после чего конкурентные решения были оценены по выбранным факторам по 10-бальной шкале путем осуществления экспертной оценки. Затем, на основании полученных значений, были посчитаны итоговые оценки конкурентоспособности. В таблице 4.1 представлена оценочная карта конкурентных технических решений.

Таблица 4.1 – Оценочная карта конкурентных технических решений

Критерии оценки	Вес критерия	Баллы				Конкурентоспособность			
		Б _ф	Б _{к1}	Б _{к2}	Б _{к3}	К _ф	К _{к1}	К _{к2}	К _{к3}
Доступность	0,25	9	2	2	2	2,25	0,5	0,5	0,5
Точность	0,2	8	9	8	7	1,6	1,8	1,6	1,4
Потребность в вычислительных ресурсах	0,2	6	4	6	8	1,2	0,8	1,2	1,6
Возможность переобучения	0,15	6	4	4	4	0,9	0,6	0,6	0,6
Универсальность	0,2	8	6	3	3	1,6	1,2	0,6	0,6
Итого	1	37	25	23	24	7,55	4,9	4,5	4,7

Исходя из проведенного анализа можно заключить, что уязвимость конкурентных технологических решений связана с доступностью, так как алгоритмы не находятся полностью в открытом доступе, а также с узкой направленностью тем обрабатываемых изображений.

4.1.2 SWOT-анализ

Была составлена матрица SWOT-анализа, представляющая сильные и слабые стороны разработки, а также потенциальные возможности и угрозы. После чего были предложены решения использования данных факторов для улучшения разработки. Матрица SWOT-анализа представлена в таблице 4.2.

Таблица 4.2 – SWOT-анализ

		Внутренние факторы	
		Сильные стороны: С1. Доступность для широкой аудитории. С2. Высокая точность. С3. Воспроизводимость.	Слабые стороны: Сл1. Потребность в вычислительных ресурсах при переобучении. Сл2. Сложность переобучения для конечного пользователя.
Внешние факторы	Возможности: В1. Расширение для обработки пакетов изображений. В2. Расширение для распознавания текста. В3. Переобучение нейронной сети на новых базах данных изображений.	В1С1С3. Реализация автоматического сохранения файла с данными о найденных текстовых областях при обработке пакета изображений. В2С1С2С3. Расширение функциональности в зависимости от нужд потребителя. В3С1С2. Повышение точности путём использования новых баз данных изображений.	В1В3Сл1. Модульное программирование частей проекта для повторного использования. В2Сл2. Переобучение нейронной сети для распознавания текста в найденных текстовых областях. В3Сл1. Реализация возможности обновления вместо повторного скачивания программы с переобученной сетью.
	Угрозы: У1. Невостребованность системы. У2. Конкуренция с аналогами.	У1С1С3. Публикация статей с описанием программы для привлечения внимания. У2С2. Отслеживание конкурентных решений и улучшение возможностей на их основании.	У1У2Сл1Сл2. Отслеживание и применение новых технологий. У2Сл1Сл2. Оптимизация переобучения для сокращения требуемых ресурсов.

Для определения комбинаций взаимосвязей областей матрицы SWOT-таблицы была составлена интерактивная матрица проекта, представленная в таблице 4.3.

Таблица 4.3 – Интерактивная матрица проекта

		Сильные стороны проекта			Слабые стороны проекта	
		С1	С2	С3	Сл1	Сл2
Возможности	B1	+		+	+	
	B2	+	+	+		+
	B3	+	+		+	
Угрозы	У1	+		+	+	+
	У2		+		+	+

В результате проведения SWOT-анализа были предложены варианты улучшения разработки на основании сильных сторон и возможностей, нейтрализации слабых сторон с помощью возможностей и нейтрализации угроз с помощью сильных сторон, а также способы избегания возможных потерь от угроз при наложении на слабые стороны. Общим направлением среди всех вариантов можно считать улучшение технологий на основании данных, полученных при наблюдении за конкурентами, или на основании возможных требований пользователей.

4.2 Планирование научно-исследовательских работ

4.2.1 Разработка графика проведения научного исследования

Для организации и систематизации работы был сформирован ряд работ и назначены исполнители для каждого этапа работы. Для реализации проекта необходимо 2 исполнителя – научный руководитель (НР), инженер (И).

Для определения трудоемкости работ использовались следующие показатели:

- Ожидаемое значение трудоемкости ($t_{ож}$),
- Продолжительность каждой работы в рабочих днях (T_p),
- Продолжительность каждой работы в календарных днях (T_k),
- Коэффициент календарности ($k_{кал}$).

Согласно производственному календарю в 2021 году 365 календарных дней, из них для 6-дневной рабочей недели 66, а для 5-дневной рабочей недели 118 выходных или праздничных дней.

Расчеты по трудоемкости выполнения работ представлены в таблице 4.4.

Таблица 4.4 - Временные показатели проведенного исследования

№	Наименование работы	Исполнители работы	Трудоемкость работ, чел.-дни			Длительность работ, дни	
			t_{min}	t_{max}	$t_{ож}$	T_p	T_k
1	Выбор направления исследования	И	1	2	1,4	1	1
2	Составление и утверждение темы	НР	1	2	1,4	1	1
		И	1	2	1,4	1	1
3	Составление календарного плана-графика	НР	1	2	1,4	1	1
		И	1	2	1,4	1	1
4	Подбор и изучение литературы	И	8	10	8,8	9	13
5	Анализ предметной области и выбор используемых методов	И	8	10	8,8	9	13
6	Проектирование программного обеспечения	И	17	25	20,2	20	30
7	Разработка программного обеспечения	И	22	35	27,2	27	40
8	Тестирование программного обеспечения и анализ полученных результатов	НР	3	4	3,4	2	2
		И	9	10	9,4	5	7
9	Согласование выполненной работы с научным руководителем	НР	2	3	2,4	1	1
		И	2	3	2,4	1	1
10	Подведение итогов, оформление работы	И	8	10	8,8	9	13
Итого		НР				5	5
		И				83	120

Диаграмма Ганта, построенная по рассчитанным значениям, представлена на рисунке 4.1.

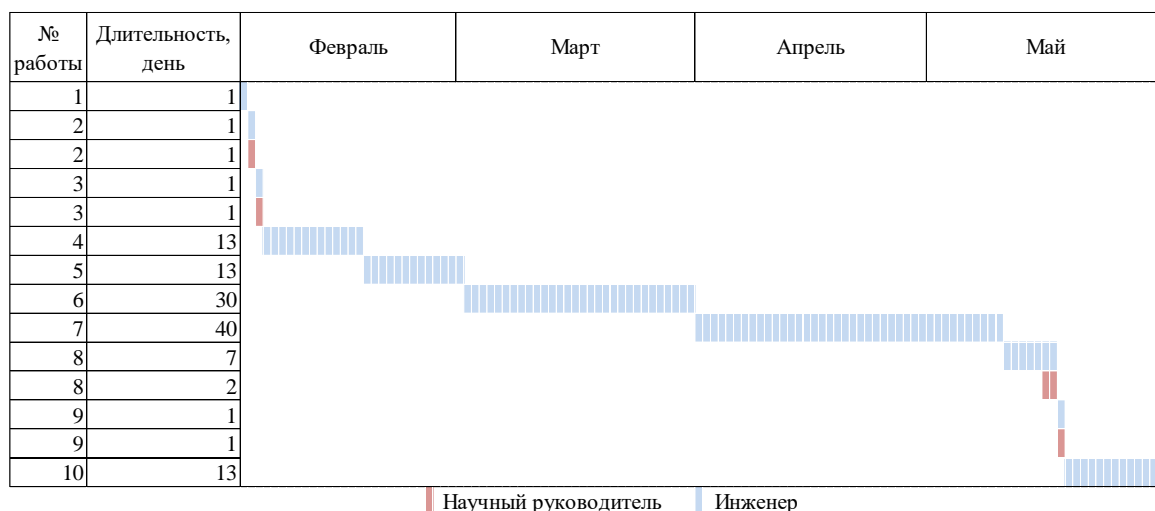


Рисунок 4.1 – Диаграмма Ганта

4.2.2 Бюджет научно-технической разработки

4.2.2.1 Расчет затрат на специальное оборудование

В данную статью затрат входят суммы, необходимые на обеспечение амортизации используемого оборудования.

Расчет амортизации персонального компьютера, используемого при написании работы: первоначальная стоимость персонального ноутбука инженера составляет 50 000 рублей; ноутбуки отнесены ко 2-ой амортизационной группе, срок полезного использования которых составляет 3 года; планируется использовать персональный компьютер для написания работы в течение 4 месяцев. Тогда:

- норма амортизации для 3 лет полезного использования:

$$A_n = \frac{1}{n} \cdot 100\% = \frac{1}{3} \cdot 100\% = 33,33 \%, \quad (4.1)$$

- годовые амортизационные отчисления для ноутбука стоимостью 50 000 рублей:

$$A_r = 50000 \cdot 0,33 = 16500 \text{ руб.} \quad (4.2)$$

- итоговая сумма амортизации основных средств для 4 месяцев использования:

$$A_m = \frac{16500}{12} * 4 = 5500 \text{ руб.}, \quad (4.3)$$

Итоговая сумма затрат на амортизацию составила 5500 руб.

4.2.2.2 Основная заработная плата исполнителей темы

В данную статью затрат входят основная заработная плата с учетом премий и доплат для исполнителей работы: инженера (студента) и научного руководителя.

Для должности доцента и степенью кандидата наук месячный оклад составляет 35 111,5 рубля, для инженера был взят оклад ассистента без ученой степени – 22 695 рублей.

Основная заработная плата от предприятия рассчитывается по следующей формуле:

$$З_{\text{осн}} = З_{\text{дн}} * Т_p, (4.4)$$

где $З_{\text{дн}}$ – среднедневная заработная плата работника, руб; $Т_p$ – продолжительность работ, выполняемых научно-техническим работником, раб. дн. (таблица 4.4).

Среднедневная заработная плата рассчитывается по формуле:

$$З_{\text{дн}} = \frac{З_m * M}{F_d}, (4.5)$$

где $З_m$ – месячный должностной оклад работника, руб.; M – количество месяцев работы без отпуска в течение года:

для 5-дневной рабочей недели $M = 11,2$ месяца;

для 6-дневной рабочей недели $M = 10,4$ месяца.

F_d – действительный годовой фонд рабочего времени научно-технического персонала, раб. дн.

В таблице 4.5 показаны количества календарных, нерабочих и праздничных дней, дней, пришедшихся на потерю рабочего времени и действительный годовой фонд рабочего времени.

Таблица 4.1 – Баланс рабочего времени

Показатели рабочего времени	Дни	
	Инженер	Научный руководитель
Календарные дни	365	365
Нерабочие дни	118	66
Потери рабочего времени	24	56
Действительный годовой фонд рабочего времени	223	243

Месячный должностной оклад работника рассчитывается по формуле:

$$З_m = З_{\text{ок}} * k_p, (4.6)$$

где $З_{\text{ок}}$ – размер оклада руководителя (указываете должность руководителя); k_p – районный коэффициент, равный 1,3 (для Томска).

Таблица 4.6 - Расчёт основной заработной платы

Исполнители	$З_{\text{ок}}$, руб.	k_p	$З_m$, руб.	$З_{\text{дн}}$, руб.	$Т_p$, раб. дн.	$З_{\text{осн}}$, руб.
Инженер	22695,00	1,3	29503,50	1481,79	83	122988,58

Научный руководитель	35111,50	1,3	45644,95	1953,53	5	9767,64
Итого:						132756,22

4.2.2.3 Дополнительная заработная плата исполнителей темы

В данную статью расходов входит заработная плата, начисленная рабочим и служащим не за фактически выполненные работы или проработанное время, а в соответствии с действующим законодательством, в том числе оплата очередных отпусков рабочих, времени, связанного с выполнением государственных и общественных обязанностей.

Зная основную заработную плату, можно рассчитать дополнительную заработную плату в размере 12-15% от основной:

$$З_{\text{доп}} = k_{\text{доп}} \cdot З_{\text{осн}}, (4.7)$$

где $k_{\text{доп}}$ – коэффициент дополнительная заработная плата (принимается равным 0,12), $З_{\text{осн}}$ – основная заработная плата.

Таблица 4.2 – Расчет дополнительной заработной платы

Исполнители	$k_{\text{доп}}$	$З_{\text{осн}}$, руб.	$З_{\text{доп}}$, руб.
Инженер	0,12	122988,58	14758,63
Научный руководитель	0,12	9767,64	1172,12
Итого			15930,75

4.2.2.4 Отчисления во внебюджетные фонды (страховые отчисления)

В данную статью затрат включаются отчисления в органы социального страхования, пенсионные фонды и отчисления медицинского страхования.

Отчисления во внебюджетные фонды рассчитываются как:

$$З_{\text{внеб}} = k_{\text{внеб}} \cdot (З_{\text{осн}} + З_{\text{доп}}), (4.8)$$

где $k_{\text{внеб}}$ – коэффициент отчислений во внебюджетные фонды, $З_{\text{осн}}$ – основная заработная плата, $З_{\text{доп}}$ – дополнительная заработная плата.

Для учреждений, осуществляющих образовательную и научную деятельность в 2014 году водится пониженная ставка – 27,1%.

Таблица 4.3 – Расчет отчислений во внебюджетные фонды

Исполнители	$k_{\text{внеб}}$	$З_{\text{доп}}$, руб.	$З_{\text{осн}}$, руб.	$З_{\text{внеб}}$, руб.
Студент	0,271	14758,63	122988,58	37329,49
Научный руководитель	0,271	1172,12	9767,64	2964,67
Итого				40294,17

4.2.2.5 Накладные расходы

Накладные расходы учитывают прочие затраты организации, не попавшие в предыдущие статьи расходов: печать и ксерокопирование материалов, оплата услуг связи, электроэнергии и т.д. Их величина определяется согласно следующей формуле:

$$З_{\text{накл}} = k_{\text{накл}} \cdot (\text{сумма статей расходов}), (4.9)$$

где $k_{\text{накл}}$ – коэффициент накладных расходов, принятый за 16%.

Таблица 4.4 – Расчет накладных расходов

Статьи затрат	Сумма, руб.
Затраты на амортизацию	5500
Затраты на основную заработную плату	132756,22
Затраты на дополнительную заработную плату	15930,75
Затраты на отчисления во внебюджетные фонды	40294,17
Накладные расходы	31116,98

4.2.2.6 Формирование бюджета затрат

Рассчитанные величины затрат научно-исследовательской работы являются основой для формирования бюджета затрат проекта. Результаты составления итогового бюджета разработки представлены в таблице 4.10.

Таблица 4.5 – Бюджет затрат на разработку

Наименование	Сумма, руб.
Затраты на амортизацию	5500
Затраты на основную заработную плату	132756,22
Затраты на дополнительную заработную плату	15930,75
Затраты на отчисления во внебюджетные фонды	40294,17
Накладные расходы	31116,98
Общий бюджет	225598,12

4.3 Определение ресурсной (ресурсосберегающей), финансовой, бюджетной, социальной и экономической эффективности исследования

Для оценки эффективности исследования сравним реализуемый вариант исполнения и один из аналогов, представленный в пункте 4.1.1. Выбирается алгоритм, описанный в статье «U-Net-bin: hacking the document image binarization contest», в котором производится детектирование рукописного текста, так как он обладает наибольшим коэффициентом конкурентоспособности среди аналогов.

4.3.1 Определение финансовой эффективности исследования

Интегральный финансовый показатель разработки определяется как:

$$I_{\Phi}^{\text{исп.}i} = \frac{\Phi_{pi}}{\Phi_{\max}}, (4.10)$$

где $I_{\Phi}^{\text{исп.}i}$ – интегральный финансовый показатель разработки; Φ_{pi} – стоимость i -го варианта исполнения; Φ_{\max} – максимальная стоимость исполнения научно-исследовательского проекта.

Ввиду отсутствия программного кода аналога в открытом доступе, а также отсутствия схематического описания структуры в статье, реализация аналога займёт большее время. Добавляем 1 месяц к времени разработки текущего исполнения, тогда стоимость увеличится примерно на 25%.

$$I_{\Phi}^p = \frac{\Phi_i^p}{\Phi_{\max}} = \frac{225598,12}{225598,12 * 1,25} = 0,8,$$

$$I_{\Phi}^1 = \frac{\Phi_i^1}{\Phi_{\max}} = \frac{225598,12 * 1,25}{225598,12 * 1,25} = 1.$$

4.3.2 Определение показателя ресурсоэффективности

Интегральный показатель ресурсоэффективности вариантов исполнения объекта исследования определяется следующим образом:

$$I_p^{\text{исп.}i} = \sum_n (a_i \cdot b_i), (4.11)$$

где I_{pi} – интегральный показатель ресурсоэффективности для i -го варианта исполнения разработки;

a_i – весовой коэффициент i -го варианта исполнения разработки;

b_i – бальная оценка i -го варианта исполнения разработки;

n – число параметров сравнения.

В таблице 4.11 представлены результаты для текущей разработки и аналога.

Таблица 4.11 - Сравнительная оценка характеристик вариантов исполнения проекта

Объект исследования Критерии	Весовой коэффициент параметра	Разработка	Аналог 1
---------------------------------	-------------------------------	------------	----------

1. Доступность	0,25	5	1
2. Точность	0,2	4	5
3. Потребность в вычислительных ресурсах	0,2	3	2
4. Возможность переобучения	0,15	3	2
5. Универсальность	0,2	4	3
Итого	1	3,9	2,55

4.3.3 Интегральный показатель эффективности

Интегральный показатель эффективности вариантов исполнения разработки определяется на основании интегрального показателя ресурсоэффективности и интегрального финансового показателя по формуле:

$$I_{\text{исп.}i} = \frac{I_{\text{р}}^{\text{исп.}i}}{I_{\text{ф}}^{\text{исп.}i}}, (4.12)$$

Сравнение интегрального показателя эффективности вариантов исполнения разработки позволит определить сравнительную эффективность проекта (таблица 4.12) и выбрать наиболее целесообразный вариант из предложенных. Сравнительная эффективность проекта рассчитывается по формуле:

$$\mathcal{E}_{\text{ср}} = \frac{I_{\text{р}}}{I_{\text{исп.}i}}, (4.13)$$

Таблица 4.12 - Сравнительная эффективность разработки

№	Показатели	Разработка	Аналог 1
1	Интегральный финансовый показатель разработки	0,8	1
2	Интегральный показатель ресурсоэффективности разработки	3,9	2,55
3	Интегральный показатель эффективности	4,875	2,55
4	Сравнительная эффективность вариантов исполнения	1,91	

Сравнение значений интегральных показателей эффективности вариантов исполнения разработки показало, что более эффективным вариантом решения технической задачи с позиции финансовой и ресурсной эффективности является вариант разработки, представленный в работе.

4.4 Вывод по разделу

В ходе данной работе был проведён анализ конкурентных решений, в результате которого в соответствии с оценочной картой конкурентных решений разрабатываемая система имеет лучшие качества. Во время проведения SWOT-анализа были выделены сильные и слабые стороны проекта, а также возможности и угрозы его реализации. Для определения комбинаций взаимосвязей областей матрицы SWOT-таблицы была составлена интерактивная матрица проекта.

В рамках процессов инициации были определены структуры и трудоемкости работ. План проекта представлен в виде диаграммы Ганта.

Был сформирован бюджет разработки, который включил в себя затраты на амортизацию, заработную плату, отчисления во внебюджетные фонды и накладные расходы. Общий бюджет проекта получился равным 225598,12 рублей.

Были рассчитаны интегральный финансовый показатель в ходе оценки бюджета затрат вариантов исполнения предлагаемой разработки и аналога, а также интегральный показатель ресурсоэффективности. На их основании рассчитан интегральный показатель эффективности, который позволил сравнить варианты исполнения и рассчитать сравнительную эффективность разработки. Более эффективным вариантом решения технической задачи с позиции финансовой и ресурсной эффективности является вариант разработки, представленный в работе.

5 Социальная ответственность

Целью выпускной квалификационной работы является разработка нейросетевого алгоритма распознавания надписей на изображениях реальных сцен.

Реализация алгоритма подразумевает собой написание сверточной нейронной сети U-Net. Для обучения сети используется находящаяся в открытом доступе база изображения KAIST. Результатом работы алгоритма для каждого изображения является маска данного изображения, на которой отмечены текстовые области.

Разработанное программное обеспечение может быть использовано в исследовательских учреждениях, предприятиях или независимыми исследователями в качестве предобработки изображений реальных сцен для дальнейшего извлечения из сегментированных областей текстовой информации.

Рабочей зоной при разработке алгоритма является рабочее место для выполнения работ сидя с использованием персональной электронно-вычислительной машины.

5.1 Правовые и организационные вопросы обеспечения безопасности

5.1.1 Специальные (характерные для проектируемой рабочей зоны) правовые нормы трудового законодательства

Основным документом, регулирующим отношения между работником и работодателем, является трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 05.04.2021). В нём содержатся основные положения об организации труда и его управлении, материальной ответственности, разрешению трудовых споров и т.д. [24].

Теоретическая разработка алгоритма относится к офисной работе, которая соответствует оптимальным условиям труда. То есть, воздействие на работника вредных и (или) опасных производственных факторов отсутствует или их уровни воздействия не превышают уровни, установленные нормативами условий труда и принятые в качестве безопасных для человека [25].

5.1.2 Организационные мероприятия при компоновке рабочей зоны

Исходя из общих принципов организации рабочего места, основными элементами рабочего места программиста являются: рабочий стол, рабочий стул (кресло), системный блок, монитор, клавиатура, мышь.

СП 2.4.3648-20 предъявляют требования к организации рабочих мест пользователей персональных электронных средств обучения (ЭСО): должна обеспечиваться зрительная дистанция до экрана не менее 50 см. Линейные размеры (диагональ) экрана ЭСО должны соответствовать гигиеническим нормативам: минимальный размер монитора персонального компьютера и ноутбука - не менее 39,6 см [26].

Рабочее место при выполнении работ сидя регламентировано ГОСТ 12.2.032-78 [27]. Конструкция рабочего места должна обеспечивать оптимальное положение работающего и обеспечивать трудовые операции в пределах зоны досягаемости моторного поля. Оптимальное положение может быть достигнуто регулированием высоты рабочей поверхности, сиденья и пространства для ног. Если невозможно осуществить регулирование высоты рабочей поверхности, допускается проектировать и изготавливать оборудование с нерегулируемыми параметрами рабочего места. Параметры для нерегулируемого рабочего места указаны в таблице 5.1 в зависимости от пола рабочего.

Таблица 5.1 - Параметры для нерегулируемого рабочего места

Параметры	Требования в соответствии с ГОСТ 12.2.032-78			Фактические показатели
	Женщины	Мужчины	Женщины и мужчины	Женщина
Высота рабочей поверхности, мм	700	750	725	760
Высота сиденья, мм	400	430	420	450

Также регламентированы требования к размещению органов управления: в данном контексте клавиатура и мышь. В соответствии с разделением рабочего пространства на зоны данные органы управления являются наиболее важными и

располагаются в зоне 1, соответствующей горизонтальной плоскости под углом $\pm 30^\circ$ от сагиттальной плоскости.

Размещение средств отображения информации касается расположения монитора, как главного средства отображения. Монитор является часто используемым средством, однако он не требует точного и быстрого считывания показаний, поэтому допускается располагать в вертикальной плоскости под углом $\pm 30^\circ$ от нормальной линии взгляда и в горизонтальной плоскости под углом $\pm 30^\circ$ от сагиттальной плоскости.

5.2 Профессиональная социальная безопасность.

5.2.1 Анализ вредных и опасных факторов, которые могут возникнуть на рабочем месте при разработке

Согласно ГОСТ12.0.003-2015 на рабочем месте при разработке могут возникнуть вредные и опасные производственные факторы [28]. Перечень возможных факторов представлен в таблице 5.2.

Таблица 5.2 – Возможные опасные и вредные факторы

Факторы (ГОСТ 12.0.003-2015)	Этапы работ			Нормативные Документы
	Разработка	Изготовле- ние	Эксплуата- ция	
1. Отсутствие или недостаток необходимого искусственного освещения	+	+	+	СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95*
2. Отклонение показателей микроклимата в закрытом помещении	+	+	+	СанПиН 2.2.4.548-96. Гигиенические требования к микроклимату производственных помещений
3. Повышенный уровень электромагнитных излучений	+	+	+	ГОСТ 12.1.006-84 ССБТ. Электромагнитные поля радиочастот. Общие требования безопасности
4. Повышенный уровень шума на рабочем месте	+	+	+	1. ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности 2. СН 2.2.4/2.1.8.562-96 Шум на рабочих местах, в помещениях жилых,

				общественных зданий и на территории жилой застройки
5. Умственное перенапряжение		+		Р 2.2.2006-05 Гигиена труда. Руководство по гигиенической оценке факторов рабочей среды и трудового процесса. Критерии и классификация условий труда.
6. Перенапряжение зрительного анализатора	+	+	+	ГОСТ Р 50948-2001 Средства отображения информации индивидуального пользования. Общие эргономические требования и требования безопасности.
7. Статические физические нагрузки, связанные с рабочей позой	+	+		Р 2.2.2006-05 Гигиена труда. Руководство по гигиенической оценке факторов рабочей среды и трудового процесса. Критерии и классификация условий труда.
8. Повышенное значение напряжения	+	+	+	1. ГОСТ 12.1.019-2017 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты 2. Правила устройства электроустановок (Седьмое издание)

5.2.1.1 Отсутствие или недостаток необходимого искусственного освещения

Освещение на рабочем месте в значительной степени влияет на восприятие получаемой информации. Даже при работе с экраном компьютера необходимо дополнительное освещение.

В качестве основного освещение было использовано искусственное освещение, которое нормируется в соответствии с СП 52.13330.2016 [29]. Данный тип освещения может применяться как индивидуально, так и совместно с естественным освещением, тогда освещение называется совмещенным.

В общеобразовательных организациях и организациях профессионального и высшего образования для компьютерных залов следует применять совмещенное освещение, нормы которого приведены в таблице 3. В соответствии с санитарными правилами характеристики зрительной работы в компьютерных залах соответствуют очень высокой точности для рабочего стола и высокой точности для экрана дисплея. Подразумевается, что в обоих случаях

относительная продолжительность зрительной работы при направлении зрения на рабочую поверхность менее 70%.

Таблица 5.3 - Нормативные показатели освещения для компьютерных залов

Тип освещения	Параметры		Рабочая поверхность	
			Экран дисплея	Рабочий стол
Искусственное	Освещенность рабочих поверхностей, лк	при комбинированном освещении	-	500/300
		при общем освещении	200	400
	Объединенный показатель дискомфорта UGR, не более		-	14
	Коэффициент пульсации освещенности, %, не более		-	5
	Индекс цветопередачи источников света R2		-	80
Естественное	КЕО еп, %	при верхнем или комбинированном освещении	-	3,5
		при боковом освещении	-	1,2
Совмещенное	КЕО еп, %	при верхнем или комбинированном освещении	-	2,1
		при боковом освещении	-	0,7

5.2.1.2 Отклонение показателей микроклимата

Микроклимат производственных помещений регламентируется в соответствии с СанПиНом 2.2.4.548-96 и характеризуется следующими параметрами [30]:

- температура;
- температура поверхностей;
- относительная влажность воздуха;
- скорость движения воздуха.

Самочувствие работника напрямую зависит от значений данных параметров. Высокая или низкая температура в комплексе с повышенной влажностью воздуха могут привести соответственно к перегреву или переохлаждению организма. При отсутствии движения воздуха в рабочем

помещении перегрев наступит намного быстрее, аналогично, высокая скорость движения воздуха ускорит переохлаждение.

В таблице 5.4 представлены оптимальные и допустимые величины показателей микроклимата на рабочих местах производственных помещений. Величины соответствуют Ia категории интенсивности энергозатрат, так как вся работа производится сидя и сопровождающиеся незначительным физическим напряжением [30].

Таблица 5.4 - Оптимальные и допустимые величины показателей микроклимата на рабочих местах производственных помещений

Величины показателей	Период года	Температура воздуха, °С	Температура поверхностей, °С	Относительная влажность воздуха, %	Скорость движения воздуха, м/с
Оптимальные	Холодный	22 – 24	21 – 25	60 – 40	0,1
	Теплый	23 – 25	22 – 26	60 – 40	0,1
Допустимые	Холодный	20 – 25	19 – 26	15 – 75	0,1
	Теплый	21 – 28	20 – 29	15 – 75	0,1 – 0,2

5.2.1.3. Повышенный уровень электромагнитных излучений

При работе с компьютерной техникой неизбежно возникновение электромагнитных полей. Излучения данных полей регламентируются в соответствии с ГОСТ 12.1.006-84 и оцениваются показателями интенсивности поля и создаваемой им энергетической нагрузкой [31].

Электромагнитные поля воздействуют на нервную систему человека, чувствительную к электрическим сигналам. Данное воздействие может вызывать повышение температуры тела, что вредно для тканей со слаборазвитой сосудистой системой или недостаточным кровообращением. Также при действии поля происходит поляризация макромолекул тканей и ориентация их параллельно электрическим силовым линиям, что может привести к изменению их свойств: нарушению функций сердечно-сосудистой системы и обмена веществ, уменьшению количества эритроцитов в крови. Степень воздействия электромагнитных полей на организм человека зависит от расстояния до

источника, диапазона частот излучения, интенсивности воздействия, продолжительности, характера и режима облучения, размера облучаемой поверхности и особенностей организма.

ГОСТ описывает требования к проведению контроля электромагнитных полей на рабочих местах и устанавливает предельные значения показателей в соответствии с таблицей 5.5.

Таблица 5.5 - Предельные значения показателей электромагнитных излучений

Параметр	Предельные значения в диапазонах частот, МГц		
	от 0,06 до 3	от 0,06 до 3	от 0,06 до 3
Напряженность электрического поля, В/м	500	300	80
Напряженность магнитного поля, А/м	50	-	-
Энергетическая нагрузка электрического поля, (В/м) ² ч	20000	7000	800
Энергетическая нагрузка магнитного поля, (А/м) ² ч	200	200	200

Сила влияния электромагнитного поля уменьшается по мере удаления от его источника, поэтому одним из вариантов снижения влияния излучения на работника является расположение источников на некотором расстоянии от рабочего места и места отдыха. Другим вариантом защиты является экранирование источника излучения и/или рабочего места.

5.2.1.4. Повышенный уровень шума на рабочем месте

Согласно ГОСТ 12.1.003-2014 шум на рабочем месте оказывает раздражающее влияние на работника, повышает его утомляемость, а при выполнении задач, требующих внимания и сосредоточенности, способен привести к росту ошибок и увеличению продолжительности выполнения задания. Длительное воздействие шума влечет тугоухость работника вплоть до его полной глухоты [32].

Вид трудовой деятельности и рабочее место в соответствии с СН 2.2.4/2.1.8.562-96 определяют предельно допустимые уровни звукового давления, уровни звука и эквивалентные уровни звука [33]. Разработка алгоритма соответствует виду трудовой деятельности, входящему в перечень

«творческая деятельность, руководящая работа с повышенными требованиями, научная деятельность, конструирование и проектирование, программирование, преподавание и обучение, врачебная деятельность»; а рабочее место соответствует рабочему месту, входящему в перечень «рабочие места в помещениях дирекции, проектно-конструкторских бюро, расчетчиков, программистов вычислительных машин, в лабораториях для теоретических работ и обработки данных, приема больных в здравпунктах». Для данного вида трудовой деятельности и рабочего места в таблице 5.6 представлены предельно допустимые значения шума.

Таблица 5.6 - Предельно допустимые уровни звукового давления, уровни звука и эквивалентные уровни звука

Уровни звукового давления, дБ, в октавных полосах со среднегеометрическими частотами, Гц									Уровни звука и эквивалентные уровни звука (дБА)
31,5	63	125	250	500	1000	2000	4000	8000	50
86	71	61	54	49	45	42	40	38	

Наиболее типичным источником шума при данном виде деятельности будет являться компьютерная техника, в частности работающая система охлаждения ПЭВМ. Для снижения негативного влияния шума на рабочем месте могут быть применены следующие средства:

- использование охлаждающих систем с меньшим уровнем шума;
- использование звукопоглощающих материалов для системного блока компьютера для смягчения вибрации.

При отсутствии возможности избегания шума во время работы необходимо организовать места отдыха на некотором расстоянии от источников шума.

5.2.1.5. Умственное перенапряжение

Одним из видов нагрузок, используемых для оценки напряженности трудового процесса, являются нагрузки интеллектуального характера. Процесс оценки факторов рабочей среды и трудового процесса описан в Р 2.2.2006-05 [34].

Для оценки выделяют следующие критерии:

- Содержание работы (степень сложности выполнения задания);
- Восприятие сигналов (информации) и их оценка;
- Распределение функций по степени сложности задания;
- Характер выполняемой работы.

В соответствии с указанными в руководстве критериями определения степени сложности решаемых задач разработка алгоритма относится к сложным задачам.

В течение рабочего дня работнику должен быть предоставлен перерыв для отдыха, в том числе для снижения умственного напряжения при выполнении задач.

5.2.1.6. Перенапряжение зрительного анализатора

Большую часть рабочего времени занимает взаимодействие с монитором компьютера, что может негативно сказаться на зрительных анализаторах. ГОСТ Р 50948-2001 описывает требования безопасности к визуальным параметрам, которые представлены в таблице 5.7 [35].

Таблица 5.7 - Требования безопасности к визуальным параметрам

Параметр	Допустимые значения
Яркость	не менее 20 кд/м
Неравномерность яркости рабочего поля экрана	не более 20%
Яркостной контраст изображения	не менее 3:1
Временная нестабильность изображения (мелькания)	не менее 60 Гц
Пространственная нестабильность изображения (дрожание)	не более $2 \cdot 10^{-4}l$, где l - проектное расстояние наблюдения, мм

Аналогично с ситуациями с умственным перенапряжением, необходимо делать перерывы в работе для отдыха зрительных анализаторов.

5.2.1.7. Статические физические нагрузки, связанные с рабочей позой

Одной из физических перегрузок является статическая нагрузка, связанная с рабочей позой. Процесс оценки тяжести трудового процесса описан в Р 2.2.2006-05 [34]. Среди основных показателей, используемых для данной оценки, наибольшие классы будут присвоены стереотипным рабочим движениям и рабочей позе. Стереотипность рабочих движений обусловлена

локальной нагрузкой с участием мышц кистей и пальцев рук при работе с клавиатурой и мышью. А высокий класс для рабочей позы обусловлен фиксированностью позы в течении более 50% рабочего времени.

Для снижения влияния нагрузки на организм необходимо делать перерывы в работе, включающие в себя смену положения тела, а также физические упражнения.

5.2.1.8. Повышенное значение напряжения

ГОСТ 12.1.019-2017 устанавливает общие требования по предотвращению опасного и вредного воздействий на персонал электрического тока, электрической дуги и электромагнитного поля. Данные воздействия могут привести к электро-травмам и заболеваниям, включая профессиональные и производственно-обусловленные заболевания [36].

Согласно пункту 1.1.13 Правил устройства электроустановок помещения без повышенной опасности поражения людей электрическим током определяются отсутствием сырости, токопроводящей пыли, токопроводящих полов, высокой температуры [37]. Данное определение соответствует помещению, в котором проводится разработка алгоритма.

Базовым принципом защиты от поражения электрическим током и электрической дугой является условие недоступности для прикосновений проводящих частей и отсутствия опасного напряжения для доступных проводящих частей.

Для обеспечения защиты от прямого прикосновения необходимо применять следующие технические способы и средства:

- основная изоляция;
- безопасное расположение токоведущих частей;
- защитное отключение.

Опасность на рабочем месте могут представлять все используемые элементы: системный блок, монитор, клавиатуры, мышь. Перед использованием оборудования необходимо убедиться в целостности проводов и вилок электропитания и в отсутствии видимых повреждений.

5.2.2. Расчет искусственного освещения

Рассчитывается система общего равномерного освещения для помещения с размерами: длина $A = 3,05$ м, ширина $B = 2,84$ м, высота $H = 2,6$ м. Высота рабочей поверхности $h_{pn} = 0,76$ м.

Как было указано в пункте 2.1.1 характеристики зрительной работы в компьютерных залах соответствуют очень высокой точности для рабочего стола при относительной продолжительности зрительной работы при направлении зрения на рабочую поверхность менее 70%. Тогда, освещенность рабочих поверхностей при общем освещении должна быть $E = 400$ лк.

Стены оклеены светлыми обоями, соответственно, коэффициент отражения стен $R_c = 30\%$, потолок белый матовый натяжной, поэтому приравниваем его к побеленному с коэффициентом отражения $R_n = 50\%$. Рабочее пространство относится к помещениям с малым выделением пыли, поэтому коэффициент запаса $k = 1,5$.

В рабочем помещении работник должен находиться в течении всего рабочего дня, а также производить работу очень высокой точности, которая, однако не имеет повышенных требований к передаче цветов. Также помещение не обладает высокими полками. В следствии всех перечисленных условий наилучшим выбором для общего освещения будут люминесцентные лампы белой цветности. Для люминесцентных ламп коэффициент неравномерности освещения берётся $Z = 1,1$.

Для нормальных помещений с хорошим отражением потолка и стен выбираются открытые двухламповые светильники типа ОД, ОДОР, ШОД, ОДО или ООД. Так как высота помещения 2,6 м, то в соответствии с наименьшей допустимой высотой подвеса над полом необходимо выбрать тип светильника ШОД. Наивыгоднейшее расположение светильников данного типа берется с критерием $\lambda = 1,1$. Расстояние светильников от перекрытия $h_c = 0,1$ м, так чтобы высота светильника над полом не превышала наименьшую допустимой высотой подвеса.

Тогда расчётная высота светильника над рабочей поверхностью вычисляется по формуле 5.1:

$$h = H - h_c - h_{\text{рп}} = 2,6 - 0,1 - 0,76 = 1,74 \text{ м. (5.1)}$$

Расстояние между светильниками вычисляется по формуле 5.2:

$$L = \lambda * h = 1,914 \text{ м. (5.2)}$$

Расстояние от крайнего ряда светильников до стены вычисляется по формуле 5.3:

$$\frac{L}{3} = 0,638 \text{ м. (5.3)}$$

Количество рядов светильников и количество светильников в ряду рассчитывается по формулам 5.4 и 5.5, учитывая, что используется тип светильника ШОД – 2-80, длина которого $l_{\text{св}} = 1,53 \text{ м}$:

$$n_{\text{ряд}} = \frac{(B - \frac{2}{3} * L)}{L} + 1 = \frac{(2,84 - \frac{2}{3} * 1,914)}{1,914} + 1 = 1,82 \approx 2; \quad (5.4)$$

$$n_{\text{св}} = \frac{(A - \frac{2}{3} * L)}{l_{\text{св}} + 0,5} = \frac{(3,05 - \frac{2}{3} * 1,914)}{1,53 + 0,5} = 0,87 \approx 1. \quad (5.5)$$

Светильники должны быть размещены в 2 ряда по 1 светильнику на ряд. План помещения и размещения на нем светильников представлены на рисунке 5.1.

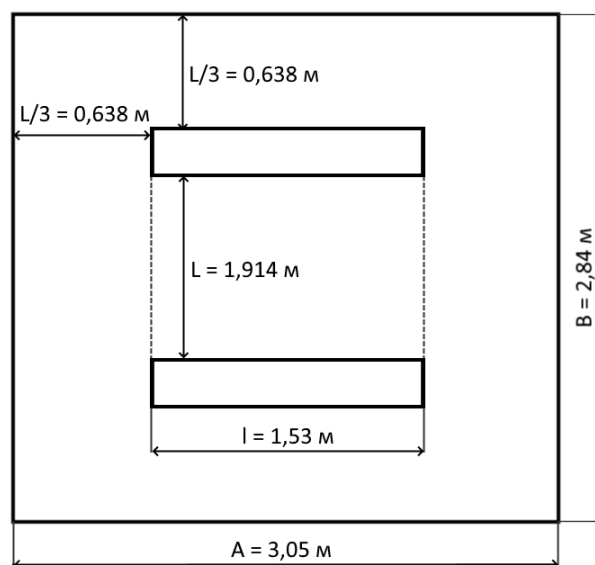


Рисунок 5.1 – План помещения и размещения светильников с люминесцентными лампами

Учитывая, что в каждом светильнике установлено две лампы, общее число ламп в помещении вычисляется по формуле 5.6:

$$N = n_{\text{ряд}} * n_{\text{св}} * 2 = 4. \quad (5.6)$$

Индекс помещения вычисляется по формуле 5.7:

$$i = \frac{S}{h * (A + B)} = \frac{3,05 * 2,84}{1,74 * (3,05 + 2,84)} = 0,845. \quad (5.7)$$

Для индекса помещения 0,845 (округление до 0,8), коэффициента отражения стен $R_c = 30\%$ и коэффициента отражения потолка $R_n = 50\%$ коэффициент использования светового потока $\eta = 0,3$. Тогда потребный световой поток ламп в каждом из рядов вычисляется по формуле 5.8:

$$\Phi = \frac{E_n * S * K_z * Z}{N_{\text{л}} * \eta} = \frac{400 * 3,05 * 2,84 * 1,5 * 1,1}{4 * 0,3} = 4764,1 \text{ лм.} \quad (5.8)$$

Для люминесцентной лампы белой цветности мощностью 80 Вт световой поток равен $\Phi_{\text{л.станд}} = 5200 \text{ лм.}$

Рассчитанный световой поток должен удовлетворять условию, указанному в формуле 5.9:

$$-10\% \leq \frac{\Phi_{\text{л.станд}} - \Phi_{\text{л.расч}}}{\Phi_{\text{л.станд}}} * 100\% \leq +20\%. \quad (5.9)$$

Необходимый поток лампы равен 8,38%, что удовлетворяет условиям.

Электрическая мощность осветительной установки вычисляется по формуле 5.10:

$$P = 4 * 80 = 320 \text{ Вт.} \quad (5.10)$$

5.3. Экологическая безопасность

5.3.1. Анализ влияния процесса разработки на окружающую среду

Главными источниками загрязнения окружающей среды является бытовой мусор, компьютерная техника, подлежащая утилизации, а также вышедшие из строя люминесцентные лампы, используемые для освещения помещения. Таким образом влияние производится преимущественно на литосферу.

Согласно Федеральному закону от 24.06.1998 N 89-ФЗ Об отходах производства и потребления выделяют пять классов опасности отходов в зависимости от степени негативного воздействия на окружающую среду [38]. Согласно Федеральному классификационному каталогу к IV классу малоопасных отходов относятся следующая техника:

- системный блок компьютера, утративший потребительские свойства;
- мониторы компьютерные плазменные/жидкокристаллические/электроннолучевые, утратившие потребительские свойства;
- клавиатура, манипулятор "мышь" с соединительными проводами, утратившие потребительские свойства.

В то время как к I классу чрезвычайно опасных отходов относятся лампы ртутные, ртутно-кварцевые, люминесцентные, утратившие потребительские свойства. Каждая такая лампа, кроме стекла и алюминия, содержит около 60 мг ртути, поэтому представляет собой опасный источник токсичных веществ.

Более точная оценка наличия вредных веществ и класса опасности производится либо специалистом на предприятии, либо сторонней специализированной организацией. Утилизация должна производиться в соответствии с этапами технологического цикла отходов, описанных в ГОСТ Р53692-2009 [39].

5.4. Безопасность в чрезвычайных ситуациях

5.4.1. Анализ вероятных ЧС, которые может инициировать объект разработки и которые могут возникнуть на рабочем месте при проведении разработки

Среди чрезвычайных ситуация, возможных для данного рабочего места можно выделить: пожар. Пожар является наиболее типичной чрезвычайной ситуацией ввиду наличия в помещении компьютерной техники, неполадки с которой могут привести его появлению. Он может возникнуть вследствие причин электрического и неэлектрического характеров. К причинам

электрического характера можно отнести короткое замыкание, искрение, статическое электричество. К причинам неэлектрического характера относится неосторожное обращение с огнём, оставление без присмотра нагревательных приборов. Пожар может привести к травме, отравлению или гибели человека, а также к материальному ущербу.

5.4.2. Обоснование мероприятий по предотвращению ЧС и разработка порядка действия в случае возникновения ЧС

Исходя из возможных причин возникновения пожара для его предотвращения могут быть применены следующие меры:

- регулярная проверка исправности работы электрических приборов, целостности соединительных проводов и вилок;
- соблюдение работниками техники безопасности при работе с электрическими приборами;
- заземления частей электрических приборов для снижения статического заряда.

В случае возникновения пожара необходимо:

- незамедлительно сообщить об это в пожарную охрану;
- принять меры по эвакуации людей;
- отключить электроэнергию, приступить к тушению пожара первичными средствами пожаротушения.

5.5 Выводы по разделу

В результате разработки раздела «Социальная ответственность» выпускной квалификационной работы магистра были изучены правовые и организационные вопросы обеспечения безопасности, изучены правовые нормы трудового законодательства и требования к компоновке рабочей зоны.

Была рассмотрена производственная безопасность, в том числе проанализированы вредные и опасные факторы, которые могут возникнуть при разработке и эксплуатации проектируемого алгоритма. Предложены

мероприятия по защите разработчика от действия опасных и вредных факторов. Был произведен расчёт искусственного освещения.

Было проанализировано влияние процесса разработки на окружающую среду и возможные чрезвычайные ситуации, которые могут возникнуть при разработке и эксплуатации проектируемого алгоритма.

Заключение

В ходе выполнения выпускной квалификационной работы был проведен анализ предметной области, включающий в себя рассмотрение существующих методов обнаружения текстовых областей на изображениях реальных сцен. В результате аналитической части в качестве классификатора была выбрана сверточная нейронная сеть архитектуры U-Net.

В качестве базы изображений был выбран набор данных KAIST Scene Text Database, для которого было проведено увеличение количества изображений за счёт применения поворотов, сжатия и разделения.

Был разработан нейросетевой алгоритм распознавания надписей на изображениях реальных сцен, для которого производился подбор таких параметров нейронной сети как размеры входных изображений, количество и типы используемых слоёв. Было произведено проектирование программного обеспечения на основании разработанного алгоритма, а также его дальнейшая реализация и тестирование.

В результате конечного обучения нейросетевого алгоритма удалось добиться высокой точности: 91,1% для обучающей выборки, 90,0% для валидационной и 88,0% для тестовой.

Было проведено сравнение с работами, проводившими тестирование на базе изображений KAIST Scene Text Database: полная набор данных и только английская часть. В ходе сравнения алгоритм данной работы показал наилучшие результаты.

Список использованных источников

1. Text line segmentation in historical document images using an adaptive U-Net architecture / O. Mechi, M. Mehri, R. Ingold, and N. Essoukri Ben Amara // 2019 International Conference on Document Analysis and Recognition (ICDAR). – 2019. – P.369-374.
2. A new U-Net based license plate enhancement model in night and day images / P. N. Chowdhury, P. Shivakumara, R. Raghavendra, U. Pal, T. Lu and M. Blumenstein // The 5th Asian Conference on Pattern Recognition. – 2019. – P.749-763.
3. Robust wide-baseline stereo from maximally stable extremal regions / J. Matas, O. Chum, M. Urban, and T. Pajdla // Image and Vision Computing. – 2004. – Vol. 22, №10. – P.761-767.
4. Neumann L. Real-time scene text localization and recognition / L. Neumann, J. Matas // 2012 IEEE Conference on Computer Vision and Pattern Recognition. – 2012. – P.3538-3545.
5. Epshtein B. Detecting text in natural scenes with stroke width transform / B. Epshtein, E. Ofek, Y. Wexler // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – 2010. – P.2963-2970.
6. Ahmed N. Discrete Cosine Transform / N. Ahmed, T. Natarajan, K. R. Rao // IEEE Transactions on Computers. – 1974. – Vol. C-23, №1. – P.90-93.
7. Zhong Y. Automatic caption localization in compressed video / Y. Zhong, H. Zhang and A. K. Jain // Pattern Analysis and Machine Intelligence, IEEE Transactions on. – 2000. – Vol. 22, №4. – P.385-392.
8. Dalal N. Histograms of Oriented Gradients for Human Detection / N. Dalal, B. Triggs // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) – 2005. – Vol. 1. – P.886-893.
9. Czarnek N. Physically Motivated Feature Development for Machine Learning Applications // Department of Electrical and Computer Engineering Duke University. – 2017.
10. Viola P. Rapid Object Detection using a Boosted Cascade of Simple Features / P. Viola, M. Jones // Proceedings of the 2001 IEEE Computer Society

Conference on Computer Vision and Pattern Recognition. – 2001. – Vol. 1. – P.511-518.

11. Ghorbel A. Generalized Haar-like filters for document analysis: application to word spotting and text extraction from comics // Document and Text Processing. Université de La Rochelle. – 2016.

12. Chen X. Detecting and reading text in natural scenes / X. Chen and A. L. Yuille // Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – 2004. – Vol. 2. – P.366-373.

13. Schmidhuber J. Deep learning in neural networks: An overview // Neural Networks. – 2015. – Vol. 61. – P.85-117.

14. Goodfellow I. J. Multi-digit number recognition from street view imagery using deep convolutional neural networks / I. J. Goodfellow, Y. Bulatov, J. Ibarz, S. Arnoud, V. Shet // Proceedings of the International Conference on Learning Representations (ICLR). – 2014.

15. Zhang, Z. Multi-oriented text detection with fully convolutional networks. / Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, X. Bai // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – P.4159-4167.

16. Ronneberger O. U-net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox // Medical Image Computing and Computer-Assisted Intervention. – 2015. – Vol. 9351. – P.234-241.

17. Bezmaternykh P.V. U-Net-bin: hacking the document image binarization contest / P.V. Bezmaternykh, D.A. Ilin, D.P. Nikolaev // Computer Optics. – 2019. – Vol. 43 №5. – P.825-832.

18. Краснящих А.В. Обработка оптических изображений. – СПб: НИУ ИТМО, 2012. – 129 с.

19. Bai B. A Seed-Based Segmentation Method for Scene Text Extraction / B. Bai, F. Yin and C. L. Liu // 2014 11th IAPR International Workshop on Document Analysis Systems. – 2014. – P.262-266.

20. Agrawal A. Enhanced characteriness for text detection in the wild / A. Agrawal, P. Mukherjee, S. Srivastava, B. Lall // Proceedings of 2nd International Conference on Computer Vision & Image Processing. – 2018. – P. 359-369.
21. Gomez L. A fast hierarchical method for multi-script and arbitrary oriented scene text extraction / L. Gomez, D. Karatzas // International Journal on Document Analysis and Recognition. – 2016. – Vol. 19, №4. – P.335-349.
22. Lee J. Page segmentation using a convolutional neural network with trainable co-occurrence features. / J. Lee, H. Hayashi, W. Ohyama, S. Uchida // 2019 International Conference on Document Analysis and Recognition (ICDAR). – 2019. – P.1023-1028.
23. Wick C. Fully convolutional neural networks for page segmentation of historical document images / C. Wick, F. Puppe // 2018 13th IAPR International Workshop on Document Analysis Systems (DAS). – 2018. – P. 287-292.
24. Трудовой кодекс Российской Федерации от 30.12.2001 N 197-ФЗ (ред. от 05.04.2021)
25. Федеральный закон от 28.12.2013 N 426-ФЗ (ред. от 30.12.2020) О специальной оценке условий труда (с изменениями и дополнениями)
26. СП 2.4.3648-20 Санитарно-эпидемиологические требования к организациям воспитания и обучения, отдыха и оздоровления детей и молодежи
27. ГОСТ 12.2.032-78 ССБТ. Рабочее место при выполнении работ сидя. Общие эргономические требования.
28. ГОСТ 12.0.003-2015 ССБТ. Опасные и вредные производственные факторы. Классификация.
29. СП 52.13330.2016 Естественное и искусственное освещение. Актуализированная редакция СНиП 23-05-95*.
30. СанПиН 2.2.4.548-96. Гигиенические требования к микроклимату производственных помещений.
31. ГОСТ 12.1.006-84 ССБТ. Электромагнитные поля радиочастот. Общие требования безопасности.
32. ГОСТ 12.1.003-2014 ССБТ. Шум. Общие требования безопасности.

33. СН 2.2.4/2.1.8.562-96 Шум на рабочих местах, в помещениях жилых, общественных зданий и на территории жилой застройки.

34. Р 2.2.2006-05 Гигиена труда. Руководство по гигиенической оценке факторов рабочей среды и трудового процесса. Критерии и классификация условий труда.

35. ГОСТ Р 50948-2001 Средства отображения информации индивидуального пользования. Общие эргономические требования и требования безопасности.

36. ГОСТ 12.1.019-2017 ССБТ. Электробезопасность. Общие требования и номенклатура видов защиты.

37. Правила устройства электроустановок (Седьмое издание).

38. Федеральный закон от 24.06.1998 N 89-ФЗ. Об отходах производства и потребления (с изменениями и дополнениями).

39. ГОСТ Р 53692-2009 Ресурсосбережение. Обращение с отходами. Этапы технологического цикла отходов.

Приложение А

Раздел 1

Text detection methods

Студент:

Группа	ФИО	Подпись	Дата
8BM92	Лобанова Виктория Александровна		

Руководитель ВКР:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИТ ИШИТР	Иванова Юлия Александровна	к.т.н.		

Консультант-лингвист отделения отделения иностранных языков ШБИП:

Должность	ФИО	Ученая степень, звание	Подпись	Дата
Доцент ОИЯ ШБИП	Айкина Татьяна Юрьевна	к.ф.н.		

1 Text detection methods

The goal of text detection and localization is detecting the presence of text and generating candidate text positions. Text regions on images can be different in size, style, orientation. Additional problems can be caused by low contrast and complex backgrounds. Existing methods for text detection and localization can be categorized into three major groups:

- connected component-based methods;
- texture-based methods;
- deep learning-based methods.

1.1 Connected component based methods

The connected component-based methods find and combine small components into a large component according to the pixels' properties, e.g. intensity, color, and stroke width, and then extract features from the connected components for classification. After the connected component extraction, these methods filter out non-text components and finally extract text from the image and combine it into text region.

The connected component-based methods have advantages of minor calculation and efficiency. However, they are plagued by several limitations such as the inability to handle rotation, scale changes, complex backgrounds, and other challenging cases.

Concerning the connected component-based methods, the most representative methods are maximally stable extremal regions (MSERs) and stroke width transform (SWT).

1.1.1 The maximally stable extremal regions

The maximally stable extremal regions (MSERs) method treats the text detection problem as an efficient sequential selection from the set of extremal regions (ERs). Extremal regions have two desirable properties [3]:

- the set is closed under the continuous one-to-one transformation of image coordinates;
- it is closed under the monotonic transformation of image intensities.

MSERs implementations for different threshold values are shown in Figure A.1.

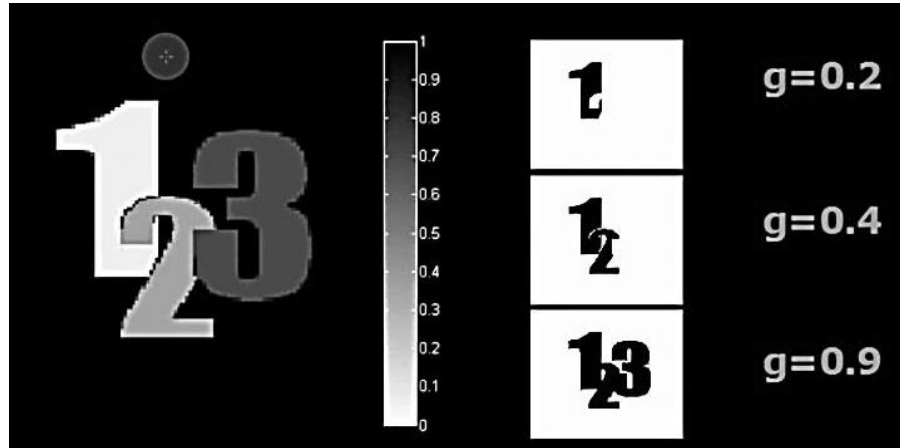


Figure A.1 - MSERs implementations

MSERs method takes up less memory, computes faster, and maintains real-time performance, but have problems with blurry images or characters with low contrast [4].

1.1.2 Stroke width transform

The primary assumption of the stroke width transform (SWT) method is that text regions in natural scenes are composed of letters that have specific geometric features and a generally consistent width over a region and also similar specific stroke length. In other words, in most cases, text regions have little stroke width variation and non-text regions have larger stroke width variation [5].

The SWT method uses the Canny edge detector to convert the image to an array containing likely stroke widths for each pixel. The next step of the algorithm is to group pixels into text regions candidates. Two neighboring pixels may be grouped if they have similar stroke width. Text tends to maintain fixed stroke width, unlike other image elements. This puts text regions apart from other non-text elements and helps to form the text line.

Stages of application of the SWT method can be found in Figure A.2.

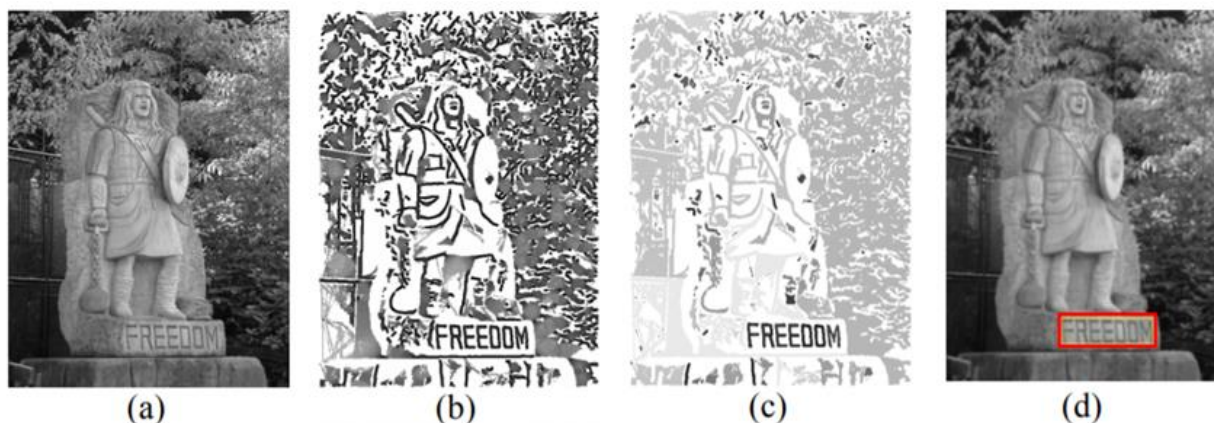


Figure A.2 - Stages of application of the stroke width transform method (a – original image, b – image with likely stroke widths for each pixel, c – image with extracting the text regions, d – image with detected text)

Experiments show that the SWT method is highly efficient for text detection. This method can detect texts in many fonts and languages, and it is insensitive to multi-scales and multi-directions. Nevertheless, the SWT method requires many human-defined constraints, so it may fail in some challenging cases.

1.2 Texture-based methods

The idea behind the texture-based method is that text in the image has distinct textural properties, which can distinguish them from the background. Generally, the features are extracted over a certain region, and a classifier is used to identify the existence of text. Texture-based methods are known to perform well even with noisy, degraded, textured, or complex texts and backgrounds. However, they are usually time-consuming as texture classification is inherently computationally dense.

The following methods were chosen for review of texture-based methods:

- the discrete cosine transform;
- the histogram of oriented gradients;
- Haar-like features.

1.2.1 The discrete cosine transform

The discrete cosine transform (DCT) method was originally intended for image compression that allows transforming the pattern space into a reduced dimensionality feature space [6].

However, the DCT method can be used to extract text regions from images. In a DCT algorithm, an image is divided into square blocks which are processed independently from each other, then the DCT of these blocks is taken, and the resulting DCT coefficients are quantized. The values of DCT coefficients reflect the local periodicity in the spatial image domain. The implicit periodicity of the DCT means that discontinuities occur at the boundaries. The algorithm proposed by Zhong uses texture features captured by the DCT coefficients of image blocks [7]. It consists of two basic steps:

- candidate regions detection in the DCT compressed domain;
- clarification of potential text regions.

Feature extraction from DCT coefficients is shown in Figure A.3.



Figure A.3 - Feature extraction from DCT coefficients (a – original image, b - DCT features from the intensity image)

The proposed algorithm is very fast because the texture features are readily available in the compressed domain and the entire post-processing is performed in reduced-resolution images. However, this algorithm is not good enough to discriminate between text and non-text regions.

1.2.2 The histogram of oriented gradients

Another texture-based method is the histogram of oriented gradients (HOG). This method is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. At first, the image is divided into small spatial regions (“cells”), then calculate histogram of gradient directions of each pixel within each cell. The concatenation of these histograms is the descriptor. The gradient image highlights the outlines and removes a lot of non-essential information [8].

Figure A.4 shows an example of histogram of oriented gradients features applied to digit recognition for digits 3 and 9. Histograms showing the prevalence of each gradient are shown for both digits [9].

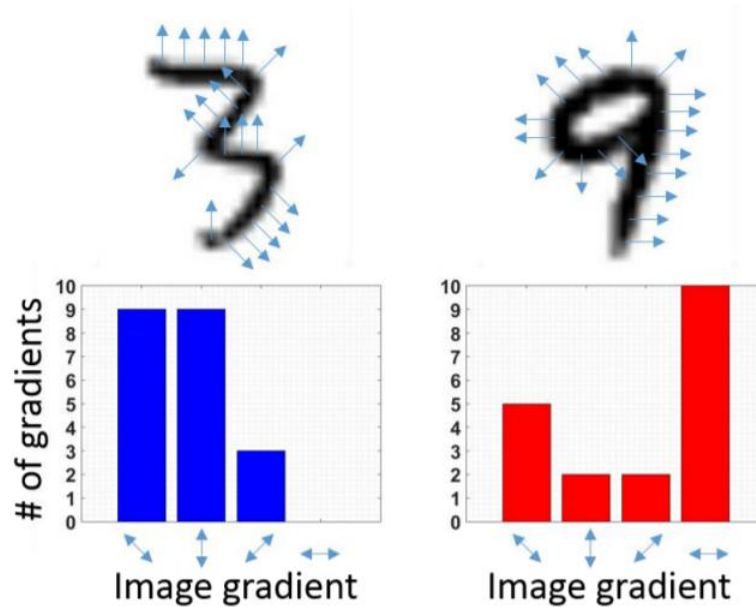


Figure A.4 - An example of histogram oriented gradient features applied to digit recognition

In the digit recognition problem, the data is easy to visualize, and readily interpretable by a human, so feature development could easily rely simply on intuition. However, many problems involve very noisy, or high dimensional, data that is not easily visualized or understood. In these instances, physical or biological models of the data can only be useful to guide how to interpret the data.

1.2.3 Haar-like features

The Haar-like features using Haar wavelets consider adjacent rectangular regions at a specific location in a detection window, sum up the pixel intensities in each region, and calculate the difference between these sums. This difference is then used to categorize subsections of an image [10].

Text, in most cases, introduces in the document some horizontal or vertical well-contrasted parts, whatever the colors of the text or the background are. Then, looking at the page, the text brings some horizontal or vertical contours. This allows estimating that Haar filters are highly recommended to detect text regions [11-12].

The examples of Haar-like features are shown in figure A.5.

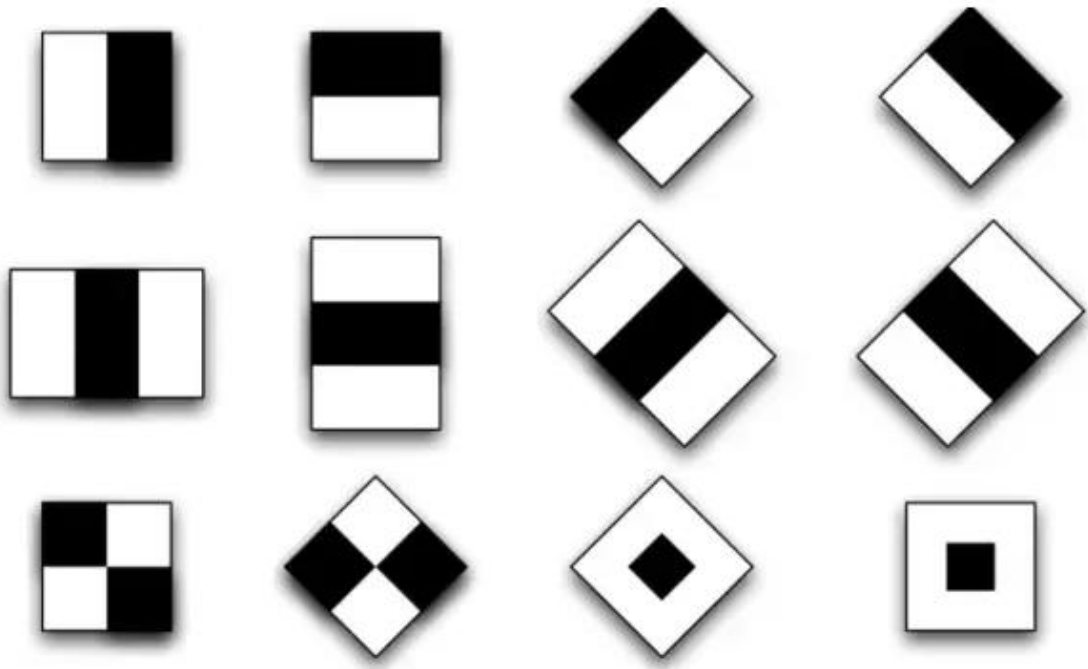


Figure A.5 - The examples of Haar-like features

The Haar-like feature problem is that the choice of the feature set is critical to the success of the algorithm. The positions of letters in the text are varied and the shapes of letters differ which can cause an additional problem for Haar-like features database preparation.

1.3 Deep learning-based methods

Deep learning-based methods mostly use the machine learning approach; the most distinctive methods are different types of neural networks.

A standard neural network (NN) consists of many simple, connected neurons, each producing a sequence of real-valued activations. Input neurons get activated through sensors perceiving the environment, other neurons get activated through weighted connections from previously active neurons. Learning is about finding weights that make the neural network exhibit desired behavior [13].

1.3.1 Convolutional neural networks

A convolutional neural network (CNN) is a class of neural networks. Like most neural networks, they contain several filtering layers that can be implemented as a discrete convolution. This makes convolutional networks computationally efficient, allowing them to scale to large images. CNN typically may include pooling layers along with traditional convolutional layers. They are used for reducing the dimensions of data by combining the outputs of neuron clusters into a single input neuron [14].

The structure of convolutional neural networks is shown in figure A.6.

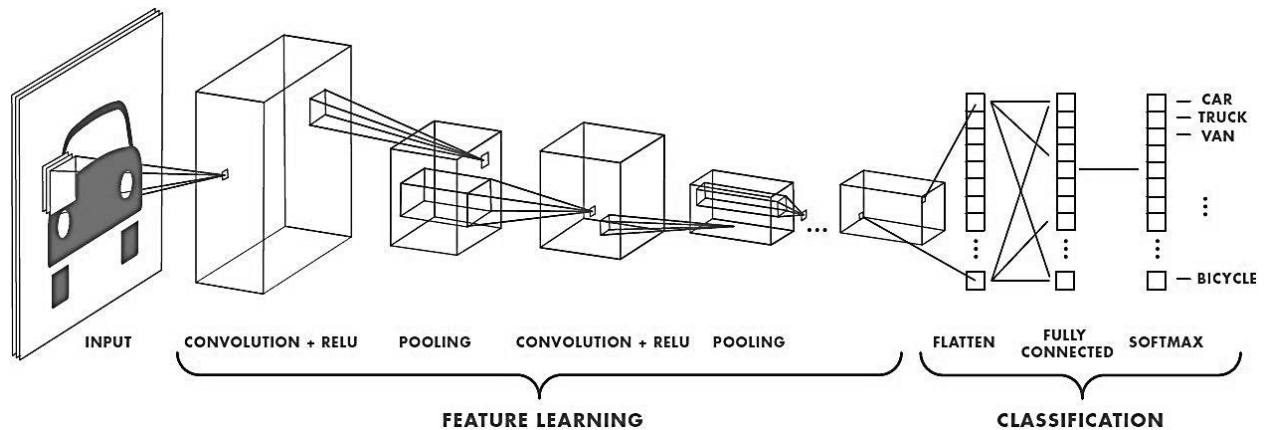


Figure A.6 - The structure of the convolutional neural network

The main advantages of CNN are being insensitive to geometric transformation, deformation, and illumination. It can extract information directly from the image with a small computational cost. The performance of text detection and recognition in the natural scene has been greatly improved due to the advanced properties of CNN. The main problem with CNN is that it only can be used for classification.

1.3.2 Fully convolutional networks

The structure of the fully convolutional network (FCN) is similar to CNN. The main difference is that the output of the FCN is the salient map of text regions generated from feature maps of each stage by the concatenation with upsampled maps. The salient map can be the preprocessing stage of localization bounding boxes of text regions [15].

The example of the structure of FCN is shown in figure A.7.

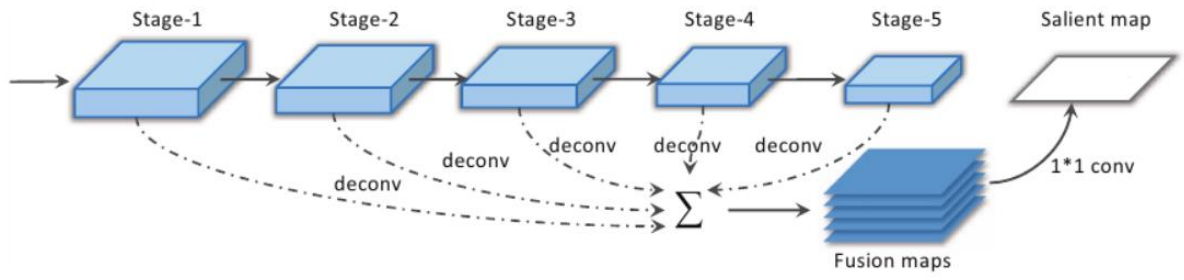


Figure A.7 - The structure of the fully convolutional network

The FCN has several failure cases: false positives and missing characters may appear in certain situations, such as extremely low contrast, curvature, strong reflect light, too closed text lines, or a tremendous gap between characters. Another limitation is the speed of the proposed method.

1.3.3 U-Net architecture

The network with U-Net architecture produces the output of the same dimensions as an input image similar to the fully convolutional neural network. One important modification in U-Net architecture is that the expansive path and the contracting path are symmetric and produce a U-shaped architecture. The expansive part consists of upsampling layers that allow having a large number of feature channels. Fully connected layers are absent in the structure, as a result, the segmentation map only contains the pixels, for which the full context is available in the input image [16].

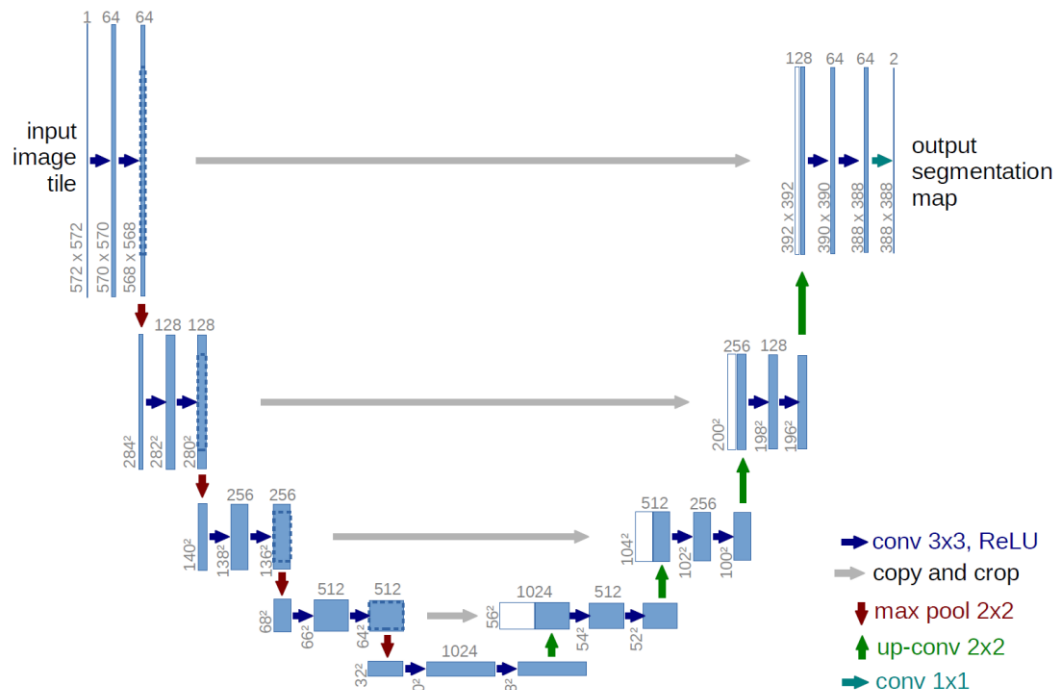


Figure A.8 - U-Net architecture

The main advantage of U-Net architecture is its ability to perform image segmentation with a small set of runs and learn with very few labeled images to capture the context in general. Also, U-Net provides pixel-wise accuracy because of same-sized input and output images [17].

1.4 Method comparison

Method comparison is shown in table A.1. It was decided to consider the following features:

- Method speed (the amount of time it takes to train a model based on this method);
- Computational cost (requirements for the equipment);
- Insensitivity to image transformation (insensitive to geometric transformation and deformation);
- Insensitivity to image noise (text variations related to noise, low contrast, and complex backgrounds);
- Method performance (accuracy or another performance measure)

Table A.1 - Method comparison

Features \ Method		Method speed	Computational cost	Insensitivity to image transformation	Insensitivity to image noise	Method performance
Connected component based methods	MSERs	Fast	Low	No	No	Medium
	SWT	Medium	Low	Yes	No	Medium
Texture-based methods	DCT	Fast	Low	No	No	Low
	Hog	Medium	Low	No	No	Low
	Haar-like feature	Medium	Low	No	No	Low
Deep learning-based methods	CNN	Slow	Low	Yes	Yes	Medium
	FCN	Slow	Medium	Yes	Yes	High
	U-Net	Slow	Medium	Yes	Yes	High

As shown in the table above, all connected components-based methods are fast, but sensitive for image deformations and have medium performance. Texture-based methods are also fast and sensitive for image deformations but have low performance. Deep learning-based methods are slow because of large training time, but insensitive for most of the image deformation and have high performance. Therefore, deep learning-based methods are more valuable than other types of methods.

CNN has fewer requirements for the equipment, but it only can be used for classification. FCN and U-Net are on the same level according to features evaluation. However, that unique architecture of the U-Net network (the expansive path is symmetric to the contracting path) provides greater possibilities for the method compared to FCN. Because of this, U-Net architecture is the best option for this research.

Приложение Б

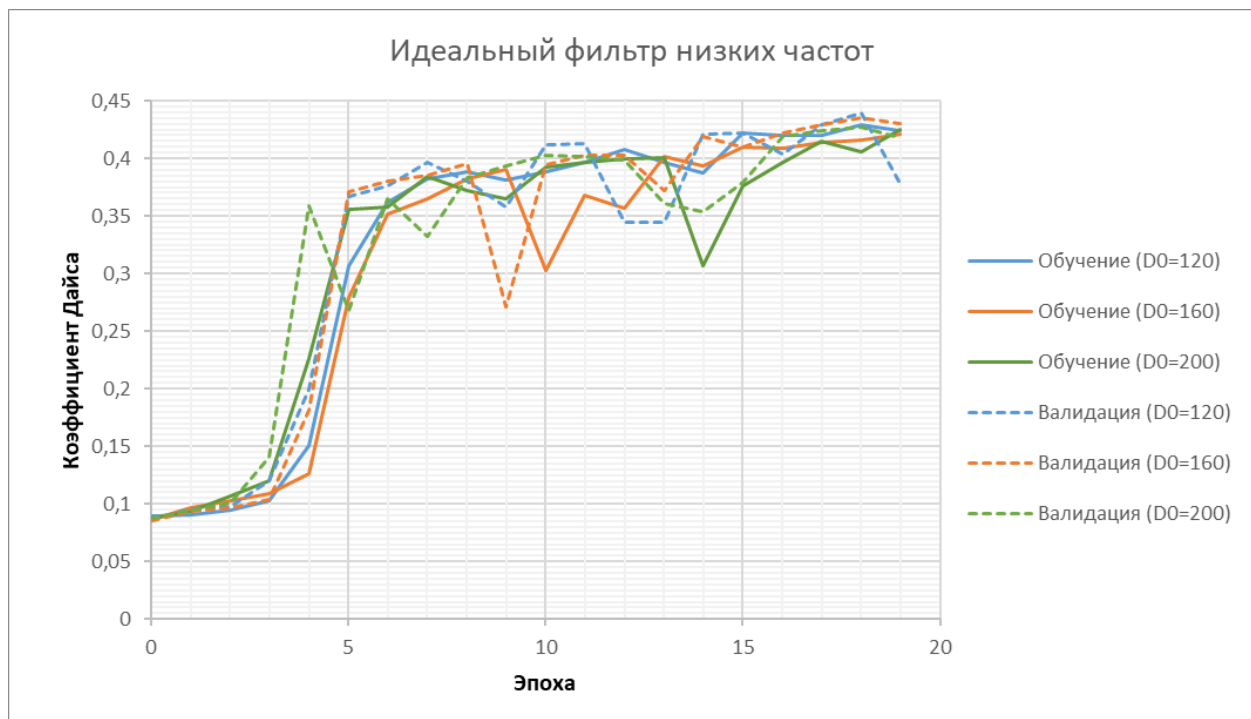


Рисунок Б.1 - Результаты обучения сети для идеального фильтра низких частот

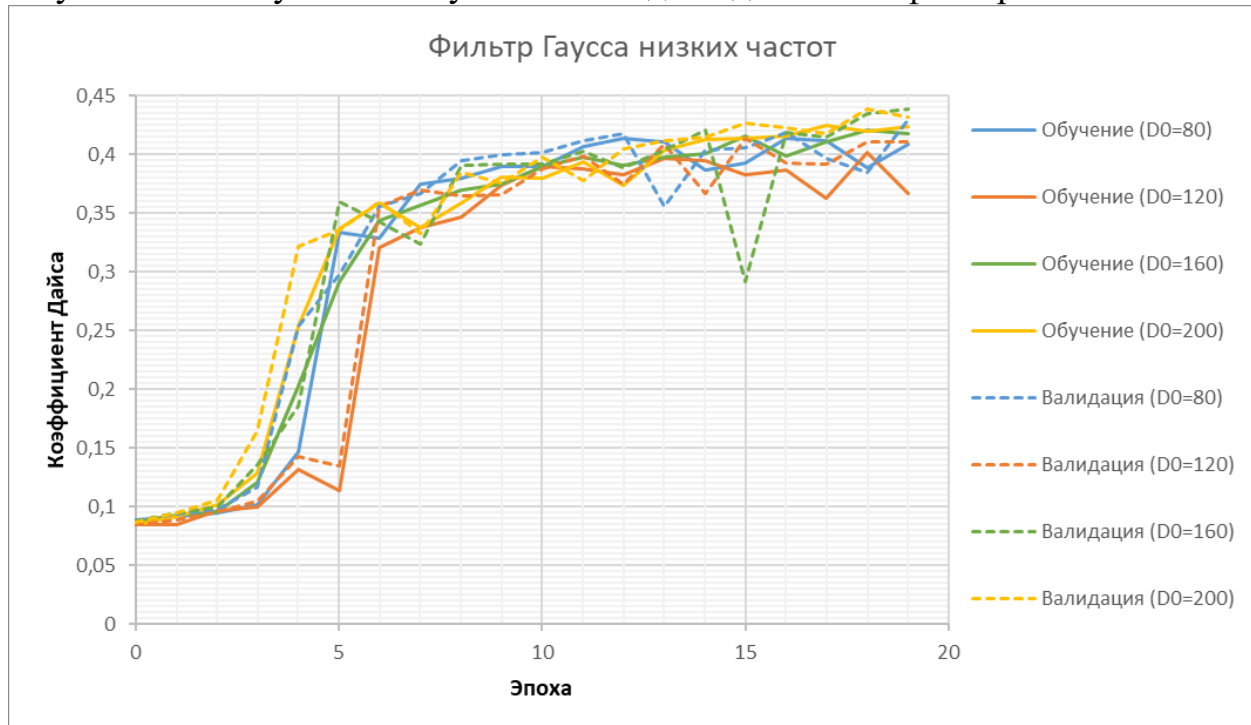


Рисунок Б.2 - Результаты обучения сети для фильтра Гаусса низких частот

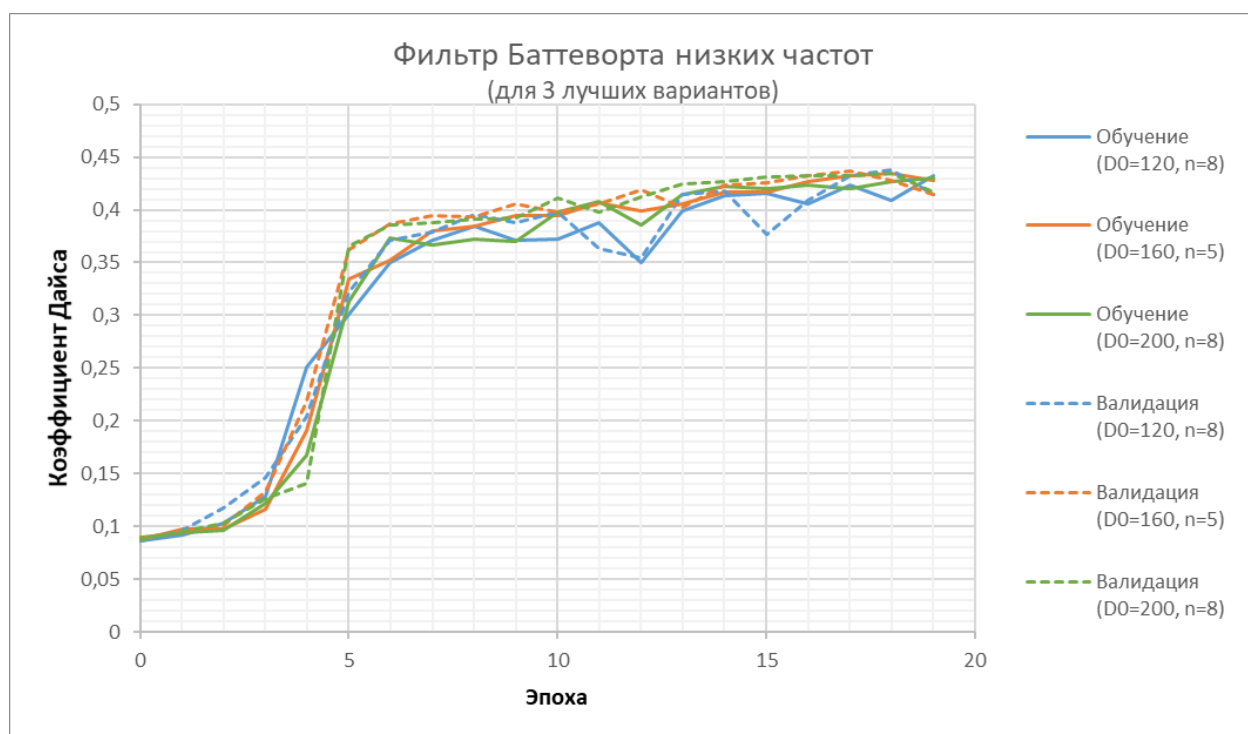


Рисунок Б.3 - Результаты обучения сети для 3 лучших вариантов фильтра Баттерворта низких частот